Distributional Changes in the Gender Wage Gap in the Post-Apartheid South African Labour Market

by

Jacqueline Nyamokami Mosomi

Thesis Presented for the Degree of

Doctor of Philosophy

In the School of Economics

UNIVERSITY OF CAPE TOWN

October 2018

Supervisor:
Professor Martin Wittenberg

DECLARATION

I declare that this thesis is my original work. Where other people's work is used, acknowledgements have been made. I declare that it has not been previously submitted for the award of a degree at any university.

Signed Signed by candidate

Candidate: Date: 16/10/2018

DEDICATION

This thesis is dedicated to my parents Mr William Mosomi and Mrs Elizabeth Mosomi, my siblings and my dearly departed grandparents.

ACKNOWLEDGEMENTS

First and foremost, I thank God for his never-ending grace that has brought me this far. I am also grateful to all the people that crossed my path during this journey and made my experience better in any way.
To my supervisor Professor Martin Wittenberg, you have been a tremendous mentor to me. Your generosity and dedication to your students is an inspiration to me. Thank you for always making time to answer my questions and for supporting my personal and professional growth. I am eternally grateful for the opportunities you opened for me by having faith in me and always pushing me out of my comfort zone.

I am highly indebted to Dr. Nicola Branson for meticulously reading through my thesis draft and offering valuable suggestions. To all my friends and colleagues at the School of Economics, this PhD journey would have been unbearable without your aid and your support. Special mention is due for Dr. David Khaoya, Takesure Bunu, Linet Arisa, Samuel Andati, Evaristo Mugocha, Dr. Francis Anguyo, Dr. Mary Karumba, Dr. Boscow Odhiambo, Dr. Adeola Oyenubi, Pinkie, Refilwe, Dr. Gibson Mudiriza, Dr. Alfred Mukong, Julie and Lerato for the stimulating discussions and the sleepless nights we would spend working together. I am also grateful to Paula Bassingthwaite and Yasmina Charles for their unfailing support and assistance.

I acknowledge, financial support from the African Economic Research Consortium (AERC) without which this research would not have been possible. I also acknowledge with thankfulness additional financial support from SALDRU through the SA-Tied project. My last year of study would have been difficult without the financial support.

My Appreciation also goes to all the participants at the MASA 2016, ESSA 2017, IUSSP 2017, CSAE 2018 IAFFE 2018 and the AERC Bi-annual conferences who gave me invaluable comments to improve my research. A special mention for Professor Tomson Ogwang' whose review of my work at the AERC Bi-annual conference helped me improve my research.

Finally, to my family, words cannot express how grateful I am for all the sacrifices you have made on my behalf. Your prayers for me was what kept me going when the journey got rough. I would also like to thank all my friends who supported me emotionally and cheered me on. Thank you for picking my calls and not letting me quit when I was ready to.

ABSTRACT

Gender inequality in the labour market is real and it matters. This thesis examines gender wage inequality in the post-apartheid South African labour market from 1993 to 2015 and finds that there is a substantial median wage gap of between 35% and 23% since the end of apartheid. This gap is unexplained by differences in human capital characteristics and it is not declining over time. There has however, been a substantial decline in the gender wage gap at the bottom of the wage distribution from 0.47 log points (about 60%) in 1997 to 0.18 log points in 2010 (about 19%). Similarly, the wage gap at the mean narrowed during the period studied, from 0.34 log points (about 40%) in 1993 to 0.15 log points (about 16%) in 2014. An examination of the gender wage gap is important because in the wake of a new democratic government in 1994 and the introduction of anti-discrimination and affirmative action policies, we expect that the gender wage gap should have narrowed over time. Interestingly, at the median where we would expect the legislation to have been binding, the gender wage gap has stagnated. We find however, that minimum wage legislation in low wage industries (agriculture and domestic work), has had some effect of narrowing the wage gap at the bottom of the distribution. This is especially interesting because this type of wage legislation was not specifically targeted at narrowing the gender wage gap.

For this analysis we utilize the Post-Apartheid Labour Market Series (PALMS) dataset from 1993-2015. To do this, we need to address several data quality issues involving how household survey data has been constructed over time. One problem we need to address is that measurement has changed over time. In chapter two of this thesis we show how measurement issues affecting data on domestic work in 1994 and 1995 led to mixed results on the gender wage gap with some researchers finding a rise in the wage gap between 1995 and 2006 and others reporting a drop depending on the choice of the base period. This analysis goes beyond pointing out the classification issues in 1994 and 1995 and takes steps to update the data and address the “missing” raw gender wage gap in these two years.

In chapter three, we analyse the trends in the gender wage gap in detail first starting with a mean decomposition using the Oaxaca decomposition method. We then extend the analysis to the entire wage distribution using the Dinardo Fortin and Lemieux (DFL) re-weighting method (DiNardo et al. 1996) for the aggregate decomposition and the Re-centred Influence Functions (RIF) method (Firpo et al. 2009) for

the detailed decomposition. Results show that the changes in the gender wage gap are heterogeneous across the wage distribution. There has been a substantial narrowing of the gender wage gap at the bottom of the wage distribution which we attribute to improved female human capital characteristics and minimum wage legislation. On the contrary the median wage gap which is greater than the mean wage gap has been stagnant and displays very little movement in the period studied. There was some decline in the gap at the 90th percentile in the period between 1993 and 2005 but the gap seems to be expanding in recent years due to a continually expanding unexplained component of the gender wage gap.

Given the snap shot nature of cross-sectional data, results from cross sectional analysis confound life cycle, generational and period effects on the gender wage gap. The fourth chapter of this thesis involves constructing synthetic cohort data from repeated cross-sections which we use to examine dynamic aspects of the gender wage gap. Life cycle trends show that younger cohorts of men and women have experienced a rise in wages over time which we attribute to improved human capital characteristics. Younger cohorts of women have on average more education than the cohorts before them and have more education than men from the same cohort. These cohorts of women are more likely to be in a skilled profession compared to women born 30 years before them and who joined the labour market during the apartheid era. Marriage rates and trade union rates have greatly declined for recent cohorts. These cohort changes mean that more recent cohorts of women have similar labour market or better labour market characteristics than men from the same cohort. This has led to the narrowing of the gender wage gap at the mean evident in the cross-sectional analysis.

The general conclusion from this analysis is that the gender wage gap persists in the South African labour market and that the experience of women at the top end of the wage distribution is different from the experience of women at the bottom end or at the median. Therefore, policies towards narrowing the gender wage gap will need to be unique to challenges women face in different parts of the wage distribution. For example, while raising the minimum wage at the bottom of the wage distribution will narrow the gender wage gap in this part of the wage distribution, narrowing the gender wage gap at the top end of the wage distribution should focus on increasing the number of women in top paying occupations. This will require among other things alleviating the disproportionate burden of care work (for example availing child care and providing creche facilities) shouldered by women to enable them to commit more time to the labour market. That non-gender specific wage legislation worked to narrow the gender wage gap at the bottom of the wage distribution is an indication that addressing the gender wage gap may require more than labour market legislation aimed at reducing gender discrimination. It may require a broader look at society and social norms that may have or may not have shifted over time.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 BACKGROUND AND MOTIVATION OF THE STUDY

1.1 INTRODUCTION

Research on inequality remains relevant both in economics and other fields of study such as sociology and psychology (Atkinson & Bourguignon 2014). This is because inequality has been found to affect welfare and life satisfaction levels of individuals in society (Clark & D'Ambrosio 2015). Inequality is undesirable because not only does it hamper poverty reduction strategies it also leads to sub-optimal allocation of resources (Okojie & Shimeles 2006). In South Africa, where according to research (Leibbrandt et al. 2010b; Bhorat et al. 2014; Wittenberg 2016b), labour market income makes up more than 85% of total household income, it is reported that individuals living in female headed households are more likely to be poor compared to individuals living in male headed households (Posel 2014; Posel & Rogan 2000; Gelb 2003). On a global scale gender wage inequality is costly to everyone; the World Bank reports that the world loses 160 trillion US dollars in human capital wealth due to gender wage inequality (Wodon & de la Brière 2018). More than two decades after the demise of the apartheid system, the question we ask is what has happened to the gender wage gap[1] in South Africa?

South Africa is an interesting case study for this research because first, South Africa's level of inequality is one of the highest in the world (Sulla & Zikhali 2018). This level of inequality has been found to be persistent and may be increasing[2] over time (Leibbrandt et al. 2010b; Sulla & Zikhali 2018). The overwhelming importance of the labour market to this high and increasing overall inequality dates back to the apartheid era. In the first half of the 20th century, South Africa was an agrarian society with black people farming both on white-owned farms as share-croppers and in the reserves (Seekings & Nattrass 2008). However, mechanization during apartheid resulted in black share-croppers being forced out of white-owned farms and into rural reserves. The rural reserves became increasingly over crowded making even peasant agriculture impossible (Nattrass & Seekings 2001; Landis 1975). Given that the Natives Land Act of 1913 prohibited black people from owning land or from making any non-labour income from

[1] In this thesis the gender wage gap is defined as $logmalewage - logfemalewage$. Additionally, we use wages and earnings interchangeably where wages are defined as labour market earnings for wage employed employees.
[2] In a recent World bank report on inequality in South Africa, Sulla & Zikhali (2018) report that the Gini coefficient (which is a common measure of inequality) rose from 0.61 in 1996 to 0.63 in 2015 after peaking at 0.64 in 2009 (see figure 1 on page XV).

land and with pass laws restricting the informal sector, wage work became the only source of income for black households (Seekings & Nattrass 2008).

The overall wage inequality in a country can affect the level of the gender wage gap (Blau & Kahn 2000, 1997). This follows from the fact that if the underlying factors behind the increase in wage inequality favour men more than women, then even with human capital characteristics or the level of discrimination staying the same, the gender wage gap will increase. An example is the rise in wage inequality due to the increase in the demand for skilled labour. The rise in the demand for skilled labour as a result of technological progress in recent years referred to as Skilled-Based Technological Change (SBTC) has been associated with the decline of blue-collar jobs in which low skilled men are over represented. This for example led to the decline of the gender wage gap at the lower end of the wage distribution in the United States in the 1980s (Blau & Kahn 1997).

In South Africa, there is evidence of skill biased growth with the continued increase in the returns to tertiary education and declining returns to incomplete secondary school education (Branson & Leibbrandt 2013). The demand for skilled workers is fuelled by the growing tertiary sectors such as finance and community services (Bhorat et al. 2014b; Bhorat 2004). Consequently, Bhorat et al. (2014b) report that there has been an increase in employment in high and medium skilled occupations such as managers, professionals and services and sales workers. On the contrary, there is evidence of a decline in the demand for unskilled workers in the primary sectors (agriculture and mining) and in the manufacturing sector (Bhorat et al. 2014b). Given that occupational segregation by gender is still evident in the South African labour market (Rospabé 2001; Gradín 2018), a decline in the share of manufacturing jobs where unskilled men are over represented is bound to have some effect on the gender wage gap.

Secondly, South Africa pre-1994 was not only characterized by racial segregation but it also displayed patriarchal attributes. Seekings & Nattrass (2008, p.82) detail how the labour and welfare system developed in the 1920s and 1930s was focussed on white households with a well-paid white male breadwinner. In this model that assumed the place of a woman to be at home, the white male bread winner was paid a “civilised” wage which was assumed to be a family wage (Seekings & Nattrass 2008). African households who were not included in the welfare system, were expected to survive on peasant agriculture and remittances from migrant labour.

Although migrant laws allowed individuals to seek jobs in the urban areas, it was mostly men who migrated partly because chiefs, husbands and fathers had the ability to prevent women and girls from migrating to towns (Posel 2004). Additionally, influx control laws and pass laws prohibited black women from migrating to towns as one could not enter the towns without accommodation. The state allowed for male dormitories in the urban areas for black men but none for black women and men and women were not allowed to stay in the same dormitory (Landis 1975). Few female teachers and nurses were able to escape the conditions in the Bantu reserves and get employment, but most women were forced into low wage jobs in white farms nearby where they were either agricultural workers or domestic workers (Landis 1975). These also happened to be the lowest paying jobs with the most precarious working conditions.

The model of the male breadwinner therefore assumed a scenario where household resources were shared amongst family members and where husbands were altruistic and would always remit money back home. This was not always the case in that there were reports of inequality within households where the patriarchal heads sort to monopolise resources which meant that rural women who could not migrate for one reason or another suffered the greatest poverty (Seekings & Nattrass 2008).

Although all women were discriminated against under the apartheid system, black women experienced discrimination in three ways. This is discrimination due to their race, their gender and their economic and social class. While white women got the right to vote in the 1930s, black men and women only obtained this right in 1994 (Seekings & Nattrass 2008). With legislation such as the colour bar, where skilled occupations were reserved for white people, combined with inferior education for black people, the apartheid system ensured that black men and women were confined to the lowest paying occupations. Additionally, pass laws and influx control laws introduced in the 1960s and 1970s, made it harder for black women to find employment in towns.

The mix of patriarchy and apartheid and their effect on the status of women in the labour market makes South Africa an important case study for understanding the evolution of the gender wage gap over time. This is because the demise of apartheid and the introduction of anti-discrimination legislation since 1994 provides us with an unnatural experiment with which to carry out this analysis. The focus of the post-apartheid government since 1994 has been to tackle all forms of inequality and discrimination through various policies and legislation (Burger & Jafta 2010). Internationally, while some studies find a positive link between equal employment legislation and labour force participation (Abe 2010), the effect of equal employment legislation on the gender wage gap is less clear (Abe 2010; Polachek 2014). Availability of 55 waves of household survey data allows us to examine how the demise of apartheid and the introduction of equal employment legislation has affected the gender wage gap in the South African labour market thus contributing to this literature.

Understanding the trend of the gender wage gap in South Africa is important in its own right, as the question of what gains women have made over time is central to the political debate. South Africa is a signatory to the Beijing Platform for Action that resulted from the fourth world conference on women in 1995 in Beijing China (Casale & Posel 2005). Additionally, the South African government has ratified many of the International Labour Organisation’s conventions that promote decent work and equality between men and women (Van Klaveren et al. 2009). An analysis of the trend of the gender wage gap is therefore one important way of understanding the status of women in South Africa.

Current results from research on the gender wage gap in South Africa are mixed with some researchers finding a rise in the gender wage gap between 1995 and 2006 and others reporting a drop. Casale (2004) found that the gender wage gap in the labour market was persistent in the period 1995 to 2001 and that even though there was an increase in the labour force participation of women, in terms of wage equality, the women remained worse off. Similarly, Ntuli (2007b) reported that in the period between 1995 and 2004, the counterfactual wage gap over the full wage distribution did not decline. Muller (2009) however, reports a gender wage gap of -0.02, 0.245, 0.209 and 0.172 log points in 1995, 1999, 2001 and 2006 respectively. Excluding 1995 from her analysis, she reports a declining gender gap between 1999 and 2006. What we will show in this thesis is that the difference in conclusions is partly due to the choice of the base period.

Another pitfall of these studies is that due to different methodologies and different analysis periods, the estimates are not directly comparable. This makes it difficult to track the size of the gender wage gap over time in South Africa. Furthermore, research on the gender wage gap in South Africa has been of a ‘snapshot’ nature (Bhorat & Goga 2013; Winter 1999; Rospabé 2001) rather than taking a longer comparative look at the evolving pattern. The disadvantage of these snapshot studies is that they do not account for inconsistencies in the data which can lead to misleading inferences. Throughout this thesis the salient theme is that for any trend analysis data quality is key and that there is value to utilising all available data instead of just a few points in time so as to separate data driven changes and real economic changes.

For the trend analysis in this thesis we utilize 55 waves of the Post-Apartheid Labour Market Series (PALMS) dataset from 1993-2015. To do this, we need to address several data quality issues involving how household survey data has been constructed over time. Since 1993, the survey instrument has undergone many changes as Statistics South Africa[3] (Stats SA) tried to improve data collection. This led to incompatibilities between surveys. In chapter two we show how measurement issues that affect data on domestic work in 1994 and 1995 lead to the raw gender wage gap being "missing" and how this has resulted in mixed results on the gender wage gap in the South African labour market. The analysis goes beyond pointing out the classification issues in 1994 and 1995 and takes steps to update the data and to address the missing raw gender wage gap in these two years. We also analyse trends in domestic work in the post-apartheid period.

Domestic work is important in its own right (ILO 2013). According to the International Labour Organization (ILO), domestic workers make up 13.6 percent of all female employees in Africa and this translates to 3,835,000 female workers (ILO 2013). A large proportion (on average 1 million) of these are employed in South Africa due to the legacy of apartheid discussed above (ILO 2013). According to Casale & Posel (2005), in 2003 one in every 4 employed women in South Africa was a domestic worker.

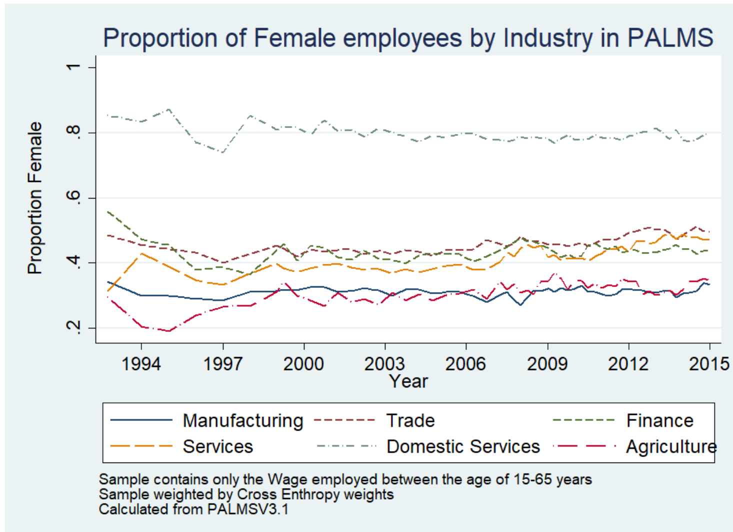
How information on domestic work is collected and documented therefore has implications for aggregate figures of employment, wage employment and female employment. The level of wages in the domestic work sector also affects the overall distribution of wages for women. In chapter two of this thesis we show how the classification of domestic workers in 1994 and 1995 led to a misleading gender wage gap in the beginning which then affected reported results on the evolution of the gender wage gap in the post-apartheid period.

The gender wage gap is partly due to occupational segregation and domestic work is almost entirely reserved for women in South Africa. A significant proportion of black and coloured women are employed in this sector which happens to be one of the lowest paying sectors (Gradín 2018). Gradín (2018) finds

[3] Statistics South Africa is the national statistical body of South Africa

that occupational segregation persists in the South African labour market and that black women and coloured women overwhelmingly fill low-paying jobs. Figure 1 shows that the proportion of women in different industries and occupations. The figure shows substantial stability of the proportions since 1994.

(a) Industry



(b) Occupation

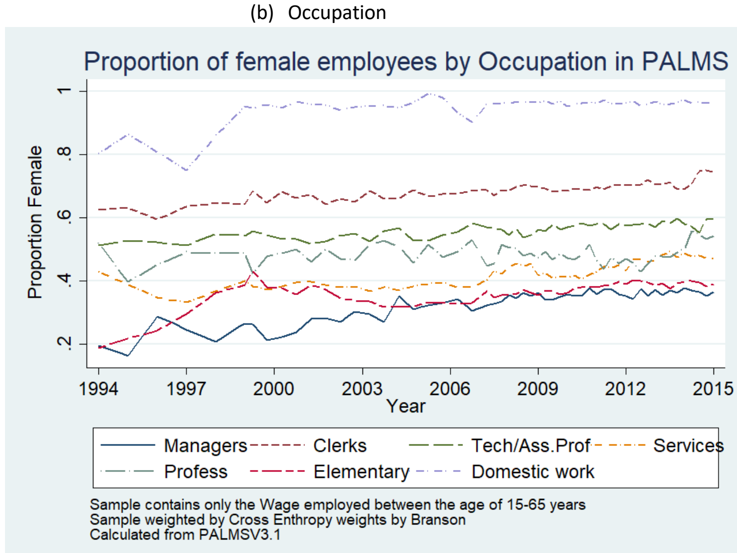


Figure 1: Employment by Industry and Occupation

It is evident that women predominantly work in four sectors; the domestic services sector (about 80% of the wage employed in this sector are women), finance (about 44%), trade (about 50%) and the services sector (about 47%). Men on the other hand are predominantly in the services, trade, finance, manufacturing and the construction sector. Looking at the main occupation variable, women are overrepresented in the lowest paying occupation (domestic work) and underrepresented in the highest paying occupations (legislators, senior officials and managers). The figure shows however, that the proportion of women in the managerial occupations has been increasing over time (19% in 1994 to 36% in 2015) almost doubling. This increase is much bigger when looked at in absolute numbers. Women managers and legislators in wage employment increased from about 58,888 in 1994 to about 279,719 in 2015. The proportion of women in technical (technical and associate professionals) and services (service workers and shop and market attendants) occupations has also been rising over time.

In chapter three of this thesis we analyse the gender wage gap in South Africa in detail and we expect that the trend of the gender wage gap will reflect the salient characteristics of the South African labour market discussed above. We start with an analysis at the mean where we compare results from a single equation estimation with a gender dummy to results from an Oaxaca decomposition (OB) (Oaxaca 1973; Blinder 1973). While OB remains the basic work horse for decomposing the gender wage gap at the mean, there is a shift towards a distributional analysis of the gender wage gap as research shows that the wage gap is not uniform over the conditional wage distribution (Albrecht et al. 2003; Chi & Li 2008; Arulampalam et al. 2007; Kee 2006).

The gender wage gap could be wider at the bottom of the wage distribution reflecting a 'sticky floor' effect, or it could be wider at the top of the distribution (a glass ceiling effect). For example, Albrecht et al. (2003), using 1998 data from Sweden, found that the gender wage gap was wider at the top of the wage distribution and therefore concluded that the gender wage gap in Sweden displayed a glass ceiling effect. Similarly, Arulampalam et al. (2007) using quantile regressions analysed the gender wage gap in 11 countries in Europe and found that in most countries the gender gap exhibited a 'glass ceiling' effect and only found a 'sticky floor' effect in two countries. In South Africa, existing studies report that the gender wage gap is wider at the bottom of the wage distribution (Ntuli 2007b; Bhorat & Goga 2013). In relation to policy, analysing the gender wage gap across the wage distribution is important because the reasons for a decline in the gender wage gap at the bottom of the wage distribution may be different from the reasons for a decline at the top, meaning that policies targeting the gender wage gap must be adjusted to account for the different experiences of women in different parts of the wage distribution.

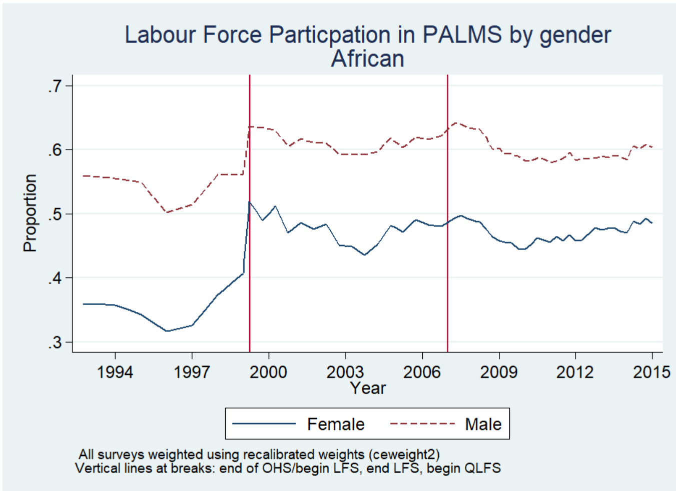
To this end, several methods that decompose the gender wage gap over the wage distribution have been suggested in the literature. These include the conditional quantile regression approach by Machado & Mata (2005), the residual imputation method by Juhn et al. (1993) and re-weighting approaches such as the Dinardo, Fortin and Lemieux re-weighting method (DiNardo et al. 1996). All these methods have been applied widely in the literature and they each have their advantages and limitations (see Fortin et al. 2011 for a review). In this thesis we utilize the re-weighting approach proposed by DiNardo et al. (1996) hereafter DFL to perform the aggregate decomposition over the conditional wage distribution and the unconditional quantile regression method proposed by Firpo et al. (2009) to perform the detailed decomposition.

According to Thornton et al. (1997), an individual's earnings are a function of macro-economic factors (inflation and general economic growth), individual productivity related to age and the birth cohort into which one was born. The birth cohort effect stems from the fact that an individual's earnings over their life cycle are affected by the macro-economic conditions when a certain cohort joins the labour market. For example, individuals that join the labour market during a recession are more likely to start at lower initial wages and an individual's wage growth is affected by the initial wages (Manning & Swaffield 2008). Similarly, individuals born into a large cohort, for example the baby boom generation, are likely to experience lower initial wages simply due to the competition that comes from a high demand for jobs (Heckman & Robb 1985; Contreras et al. 2005; Welch 1979; Easterlin 1968). Therefore, a cross-sectional analysis of the gender wage gap confounds how macro-economic conditions (period effects) affect earnings, how experience and ageing (age effects) affects earnings and how an individual's birth year (cohort effects) affects earnings. A longitudinal approach to the gender wage gap is required to account for the effect of ageing, macro-economic conditions and the birth cohort one was born into. The fourth chapter analysis constructs synthetic cohort data (quasi longitudinal data) from repeated cross-sections which we use to examine dynamic aspects of the gender wage gap.

The cohort analysis of the gender wage is important because the factors that are said to explain the presence of the gender wage gap in the labour market such as education and labour force experience are cohort sensitive. Different generations go through a different education system and labour market conditions. For example, cohorts of women joining the labour market after the introduction of equal employment legislation may have faced a more egalitarian labour market. Women from the same cohort are also likely to have undergone the same cultural socialization in terms of labour market participation or investment in human capital skills and education.

Similarly, central to the gender wage gap debate is the effect of age on the gender wage gap. Studies find that the gender wage gap is mostly smaller at younger ages soon after school completion and increases with age (Goldin 2014; Polachek 2006, 2014; Manning & Swaffield 2008). The increase is usually attributed to women dropping out of the labour market to get married and raise children. The gender wage gap, like earnings, has a life cycle profile. Over time however, marriage patterns have changed, and women are having children much later in life. The effect of age on the gender wage gap may therefore have changed over time. Beaudry & Lemieux (1999) analysed labour force participation in Canada and report that age-participation profiles of younger cohorts of women are increasingly flat, resembling those of men. This means that recent cohorts of women have a continuous attachment to the labour market.

(a)



(b)

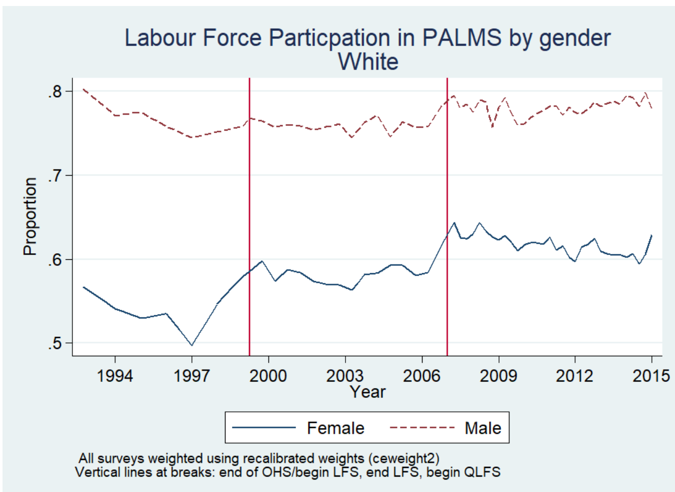


Figure 2: Labour Force Participation over time

In the South African context, several changes have taken place in the recent past that may make the experience of more recently born women different from women born before them. For example, figure two shows that even though there is still a substantial gap in labour force participation[4] between men and women, there was an increase in labour force participation in the first decade since the demise of apartheid. This increase is however greatest for African women. The labour force participation rate of African women increased from about 0.36 in 1993 to about 0.49 in 2015 after peaking at 0.51 in the year 2000. This increase in labour force participation especially among African women has been attributed to *inter alia*, improved human capital characteristics (Casale & Posel 2002), affirmative action and anti-discrimination legislation (Burger & Jafta 2010), changes in women's behavioural response towards the labour market and changes in social norms (Ntuli, 2007a; Ntuli & Wittenberg, 2013).

There is recent literature which shows that some of this increase, especially between 1998 and 2000, is due to improved data collection by statistics South Africa (Stats SA) (Casale et al. 2004; Yu 2007) and especially the improved capturing of informal activity in South Africa (Casale et al. 2004). However, starting with the participation rates in 1993, there is no denying that there was a substantial increase in female labour force participation after 1993.

Studying the mean wages of cohorts can tell us how wage inequality between and within cohorts, changes over time. For example, examining mean wages of the cohort that is 30 years old in 2000 and following this cohort for 14 years until it is 44 years old in 2014, can tell us whether the gender wage gap within this cohort increased or decreased over the lifecycle. Similarly comparing the gender wage gap of different cohorts at similar ages can tell us whether the wage gap declined or increased over time. We expect to see younger cohorts of women having wage distributions that are closer to male wage distributions. We expect the wage trajectories of younger cohorts of women to be above those of older cohorts of women given equal employment legislation, affirmative action and better education. Consequently, the wage gap should reduce for younger cohorts of women.

1.2 OBJECTIVES OF AND RATIONALE FOR THE STUDY

The above discussion leads us to three key research focus areas:

[4] Here we use the strict definition of labour force participation where an individual is said to be participating in the labour market if they are either employed or unemployed but searching.

1. What are the implications of the inconsistent measurement of domestic work in PALMS for female employment and the gender wage gap?
2. What are the most important factors in explaining the differences in wages between men and women in South Africa? Has the gender wage gap declined or increased in South Africa in the post-apartheid period? What factors can explain the changes in the gender wage gap across the wage distribution and over time?
3. Is there evidence of birth cohort effects on the gender wage gap?

1.3 THE DATA AND DATA QUALITY ISSUES

All analyses in this thesis utilise the Post-Apartheid Labour Market Series (PALMS) 1993-2015[5] (Kerr et al. 2016) dataset comprising 55 waves of South African labour market surveys stacked together. These surveys are: the 1993 Project for Statistics on Living Standards and Development (PSLSD) conducted by the Southern Africa Labour and Development Research Unit (SALDRU), the October Household Surveys (OHSs) that were started in 1993[6] and were collected annually until 1999, the Labour Force Surveys (LFSs) that were collected biannually from 2000 to 2007 and the Quarterly Labour force Surveys (QLFSs) collected from 2008.

The PALMS dataset contains the longest running series of data on the post-apartheid labour market and is therefore particularly well suited for analysing labour market outcomes. To ensure comparability, Kerr et al. (2016, 2013) have gone to great lengths to harmonize the PALMS dataset variables over time. For example, the main variable of interest in this thesis is earnings. This question was asked in all the surveys since 1994. However, the question has been changed over time (Wittenberg & Pirouz 2013). For instance, the earnings question was collapsed from 2 questions in the OHSs to one question in the LFS series (Wittenberg & Pirouz 2013). This created a break between the OHSs and the LFSs. In the QLFSs, the question reverted to two questions but is now separated into self-employment and formal employment, leaving no option for respondents to answer the two questions at the same time. The *realearnings* variable which is used for this analysis has been harmonised across the different surveys

[5] Although we have 55 waves of data (1993-2015), our data does not include wage information for 2008, 2009 and 2015. Therefore, any analysis on wages is up to 2014. Descriptive statistics will however include statistics for 2015.
[6] The 1993 OHS is however not included in PALMS as it did not cover the whole country. It excluded the former Transkei, Bophuthatswana, Venda and Ciskei (TBVC states).

and has been deflated to June 2000 using South Africa's Consumer Price Index. Additionally, the data set comes with re-calibrated weights using a cross entropy (CE) approach (Branson & Wittenberg 2014) to increase continuity and comparability between surveys over time.

Still, there remain data quality challenges which make interpretation of this data difficult (Wittenberg 2014). The data quality issues include issues regarding sampling practice and coverage, fieldwork practice, changes in the measurement instrument, choice of baseline and imputations (Wittenberg & Pirouz 2013; Wittenberg 2016b; Branson & Wittenberg 2014; Wittenberg 2014; Yu 2007; Burger & Yu 2007; Posel & Casale 2006; Casale et al. 2004; Posel & Casale 2001; Machemedze et al. 2014; Kerr & Wittenberg 2015). In the next section, we discuss some of these data quality issues and especially how they relate to the analysis of the trends in the gender wage gap in the South African labour market.

1.3.1 Changes in the Measurement Instrument

In a bid to improve data collection and reporting, Statistics South Africa (Stats SA) kept changing the measurement instrument over time. In the initial OHS, the questionnaires and sampling methodology used for the OHSs kept changing. Although this is positive in terms of improved cross-sectional data produced, it nevertheless complicates the comparability of results from different studies. For instance, increased female labour force participation between 1998 and 2000 (see figure 2) has been partly attributed to improved capturing of participation rather than a real shift in the labour force participation rate (Yu 2008). Studies show that a substantial proportion of employment in this period came from self-employed agricultural workers (subsistence farmers) (Neyens & Wittenberg 2016; Posel & Casale 2001). These researchers attribute this to the change in the definition of work[7] and further probing about the informal sector by field workers. Indeed, excluding self-employed agricultural workers results in a smoother employment series over time, although the jump in employment in this period is still significant (see figure 4b). In this thesis we exclude all the self-employed workers and focus only on the wage employed.

Related to this is the classification of domestic workers. We show in chapter two that in the OHS 1994 domestic workers were classified as self-employed elementary workers, in the OHS 1995 they were

[7] There was a change in the definition of work with the shift from the OHSs to the LFSs. Field workers were instructed to classify as employed anyone who was engaged in any informal or small-scale agricultural work even if for only an hour in the previous week (Neyens & Wittenberg 2016).

classified as self-employed domestic workers and for the rest of the surveys they were classified as wage employees. This is of particular interest to this study because the inconsistency in the classification led to a misleading gender wage gap in 1994 and 1995 which in turn led to a misleading analysis of the trend over time (see figure 3).

Figure 3 shows mean wages by gender and shows that there was practically no raw gender wage gap in 1994 and 1995. There seems to be a very high gap in 1993 which disappears in 1994 and 1995 and then resurfaces afterwards maintaining a uniform trend. This is not plausible especially given that in 1994 South Africa had just come out of the discriminatory regime of apartheid.

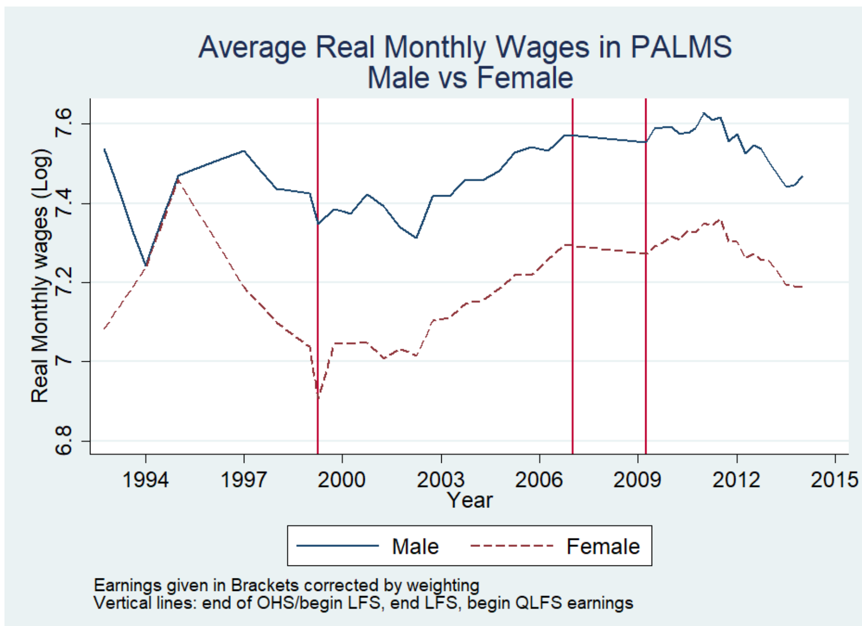


Figure 3: The Raw Gender Wage Gap in PALMS

Source: Own calculation from PALMS dataset v3.1

1.3.2 Undersampling and Coverage of Low-Income Workers in the Early OHSs

Several researchers have pointed to the issue of under sampling of low-income workers in the early October Household Surveys (Kerr & Wittenberg 2015; Wittenberg & Pirouz 2013). The 1995 OHS is reported to have under sampled small households and occupations such as mining, domestic work and subsistence agriculture where Africans are most likely to feature (Wittenberg & Pirouz 2013; Kerr & Wittenberg 2015). Kerr & Wittenberg (2015) also point to the OHS 1995 having peculiar patterns including an abnormally high level of male employment. According to Kerr & Wittenberg (2015) the under sampling was because fieldwork practices, such as the decision on the number of households to interview in a dwelling place, were left in the hands of fieldworkers. In the period prior to 1999, there was no master sample and therefore in dwelling places where there were several households, the fieldworkers only picked one household and bigger households had a higher probability of being chosen. The authors however report that there seems to be no sign of up weighting of the smaller households to compensate for the sampling strategy. In this scenario, the most likely workers to be skipped were mine workers renting lodges on top of shop buildings or hostels and live-in domestic workers who were likely to be staying in backyard extensions on their employer's compound (Machemedze et al. 2014). The issue of under sampling is important to this analysis in that we can link the under sampling to the absence of a gender wage gap in 1994 and 1995 (see figure 3). Under coverage and under sampling is also related to the sudden increase in the number of women participating in the labour market in the early 2000s.

1.3.3 Choice of baseline

Most analyses on inequality in South Africa have used household surveys from Statistics South Africa. For most studies, the OHS 1995 has been the baseline dataset used (see Branson & Wittenberg 2007 Appendix 1 for a list of studies). Branson & Wittenberg (2007) pointed out that the OHS 1995 would not make a good baseline for employment analysis as it had atypical levels of African male employment. Recently, Kerr & Wittenberg (2015) point to the OHS 1995 having peculiar patterns including few small households which means that some occupations are under sampled. A critical look at figure 3 makes clear why snap shot type of analyses may be a problem. It also shows how the choice of baseline may affect results on the gender wage gap or any analysis on wages. Studies that analysed the gender wage gap using the OHS 1994 or the OHS 1995 found either a gender wage gap in favour of African women (Hinks 2002; Grün 2004) or an insignificant gender wage gap (Winter 1999) leading these authors to note that the results seemed peculiar.

Studies that compared other years to 1995 concluded that the gender wage gap had not declined over time (Ntuli 2007b). We now know that the OHS 1995 is a misleading baseline for any analysis of wages and wage inequality due to the under sampling of low-income earning women in the early OHSs (Machemedze et al. 2014; Kerr & Wittenberg 2015). In chapter two, we discuss how fieldwork practice led to under sampling of live-in domestic workers in the 1990s and how a change in the classification of domestic workers from wage employees to self-employed elementary workers led to a non-existent gender wage gap in 1994 and 1995. Throughout this thesis we avoid the disadvantages of a snap shot data by using all 55 waves of PALMS data. This allows us to identify trends that are due to social and economic effects as opposed to data quality effects. Unlike most of the studies that base their analyses on the OHS 1995, we extend the period of analysis to 1993 by including the 1993 Project for Statistics and Living Standards (PSLSD) dataset. The inclusion of the PSLSD 1993 provides a better starting point and acts as a point of comparison for later Statistics South Africa surveys.

1.3.4 Missing and Bracket Earnings and Outliers

Another issue is that for most of the earlier papers it is not clear how authors dealt with missing earnings data, incomes given in brackets and outliers which are prevalent in some waves of the household surveys such the OHS 1999 and the LFS 2000 (Wittenberg & Pirouz 2013).

The exception is Muller (2009) who uses the sequential multiple regression imputation (SMRI) techniques to impute values for missing wage information in both the 2001 and the 2006 LFS data. The author reports that there was a smaller decline in the wage gap for part-time workers between 0.124 and 0.139 log points with imputed data than when missing values are ignored, where the decline in the wage gap was between 0.149 and 0.152 log points. For full time workers, the decline in the wage gap is larger after imputation, ranging between 0.044 and 0.051 log points, than when excluding the missing observations where the decline ranges between 0.037 and 0.047 log points (Muller 2009, p.15).

On the other hand, Bhorat & Goga (2013) use the midpoint method to account for earnings given in brackets. This method of imputing for missing earnings has been used in previous research (Posel & Casale 2006). Posel & Casale (2006) report that estimates from imputation using the midpoint method are not biased however Wittenberg (2016b) has shown that this method of imputation seems to lead to earning estimates that overstate measures of inequality. By using the earnings provided in the PALMS dataset we deal with the issue of earnings given in brackets since PALMS contains bracket-weights to account for earnings given in brackets. It also includes an outlier variable that flags outliers in the dataset (see Wittenberg 2016b for more details).

1.4 STRUCTURE OF THE THESIS

This thesis is divided into five chapters. Chapter two examines the implications of the inconsistent measurement of domestic work in the South African labour market series for female employment and the gender wage gap. A probit model is also used to understand the determinants of domestic work in the South African labour market.

Chapter three examines distributional changes in the gender wage gap in the period 1993-2014 taking these corrections into account. We combine Ordinary least squares (OLS) regression with Oaxaca decomposition to analyse the gender wage gap at the mean and then we utilise the semi parametric re-weighting method by DiNardo, Fortin and Lemieux (DFL) (DiNardo et al. 1996) and the RIF methodology (Firpo et al. 2009) to analyse the gender wage gap across the wage distribution.

We study the age (lifecycle), cohort (generational) and period effects on earnings and the gender wage gap in chapter four. This is carried out by constructing birth cohort data using repeated cross sections. The decomposition of life cycle, generational and period effects is carried out using the Deaton (1985) methodology.

In chapter five we synthesise the results from chapter two, three and four and offer suggestions for future research. Important to note is that although this thesis is under one general theme (Distributional changes in the gender wage gap in South Africa) each empirical chapter (chapter two, three and four) is presented as a stand-alone unit complete with its own introduction, literature review, methods section, results section and a discussion.

1.5 SUMMARY

This thesis examines the trend of the gender wage gap in the post-apartheid South African labour market covering the period between 1993 and 2015. The investigation is conducted using the Post-Apartheid Labour Market Series (PALMS). We find that regardless of the method used, the gender wage gap persists but that there has been a decline at the lower end of the wage distribution.

What we will show is that explicit gender based affirmative action has not been as successful as one might have thought but that other interventions such as minimum wage legislation in low income industries have worked. What this implies is that there is no blanket solution for dealing with gender wage inequality in the labour market. Women in different parts of the wage distribution face different challenges. While for women at the bottom of the wage distribution the challenge is the very low pay in the few occupations women are concentrated in such as domestic work, for women at the top of the wage distribution, the challenge is that due to child care constraints, they are unable to compete for promotions. To reduce the gender wage gap what this suggests for the literature is that legal interventions without an evaluation of social norms is not going to be enough.

A cohort analysis of the gender wage gap shows that the gender wage gap has narrowed for the most recent cohorts and it is tending to zero. This can partially explain the narrowing of the gender wage gap in recent years evident in the cross-sectional analysis. Our results show that more recent cohorts of women have better human capital characteristics than their predecessors which has led to the narrowing of the wage gap. Additionally, women have had more relative wage gains compared to their male counterparts which is another reason for the narrowing of the wage gap.

Finally, throughout this thesis, we show that, owing to changes in the manner in which household surveys have been conducted over the years, for any trend analysis, paying attention to the data quality is important. We provide several methods of dealing with data quality issues. For example, utilising all available data instead of picking a few points in time allows one to identify breaks in the data. Also, we show that there are available methods for dealing with missing data and finally, constructing cohort data is a useful way to analyse changes over time.

2 MEASUREMENT ISSUES IN THE SOUTH AFRICAN LABOUR MARKET SERIES: AN ILLUSTRATION OF THE IMPLICATIONS FOR ANALYSIS OF FEMALE EMPLOYMENT AND DOMESTIC WORK

2.1 INTRODUCTION

Data[8] from national household surveys show that domestic workers constitute a significant proportion of the employed population in South Africa (about 8 percent since 2008) and they constitute an even bigger proportion of wage employment among women (about 18 percent in 2010 and about 16 percent in 2015). Budlender (2011, p.33) using LFS 2007 and using the narrow definition of employment reports that "for female workers, the employment rate would be 6 percentage points lower and the unemployment rate 12 percentage points higher without domestic work". The domestic services sector therefore, makes a significant contribution to employment in a high unemployment[9] environment. Additionally, there is a significant proportion of South African households that depend on income from this sector which makes it an important sector in its own right.

Descriptive analysis of household surveys since 1993 reveals two trends. The first trend, which is well documented in the literature, is the "sudden" increase in female labour force participation that took place in the early 2000s and was at the time termed the "feminisation" of the labour market (Casale & Posel 2002; Casale 2004). This trend received much attention from researchers because it happened within a very short period[10] (see figure 2). Researchers have attributed this increase in labour force participation partly to changes in household formation (decline in marriage rates and migration) and partly to improved data collection especially in the informal sector and in particular the agricultural and the domestic services sectors (Casale et al. 2004; Posel & Casale 2001).

The second trend evident in descriptive statistics and which to the best of our knowledge has not been interrogated in the literature is the lack of a raw gender wage gap in 1994 and 1995 among wage employees. The gender wage gap in the post-apartheid labour market is well documented (Rospabé

[8] The statistics are the author's own calculations from the PALMS dataset. Domestic workers as a proportion of the employed population has remained about 8% since 2008.
[9] Unemployment was at 26.6% in the second quarter of 2016 (http://www.statssa.gov.za/)
[10] Branson & Wittenberg (2007) show that it was concentrated between 1998 and 2000

2001; Winter 1999; Casale 2004; Grün 2004). Gender and racial inequality studies in South Africa have attracted both local and international researchers because of South Africa's history of the apartheid regime in which discrimination was enshrined in the law. It is therefore surprising that more attention has not been paid to the non-existent gender wage gap evident in 1994 and 1995. The data shows a break in the series with a substantial raw wage gap in 1993 and then a consistent wage gap in the subsequent years after 1995 (see figure 3).

In this chapter we examine trends of employment of women in South Africa. Although a lot of work has been done on women's employment (Casale 2004; Casale & Posel 2002; Ntuli 2007a) we will show that domestic workers are key to understanding these trends and inconsistencies in the documenting of this group has led to a misleading trend of the gender wage gap over time. We show that the lack of a gender wage gap in 1994 and 1995 is related to how domestic workers were classified in these two years. Domestic workers as low-income workers contribute a significant proportion of the women in wage employment and therefore excluding many of them is going to bias downwards estimates of totals and bias upwards estimates of averages such as average wages leading to trends such as the one in figure 3.

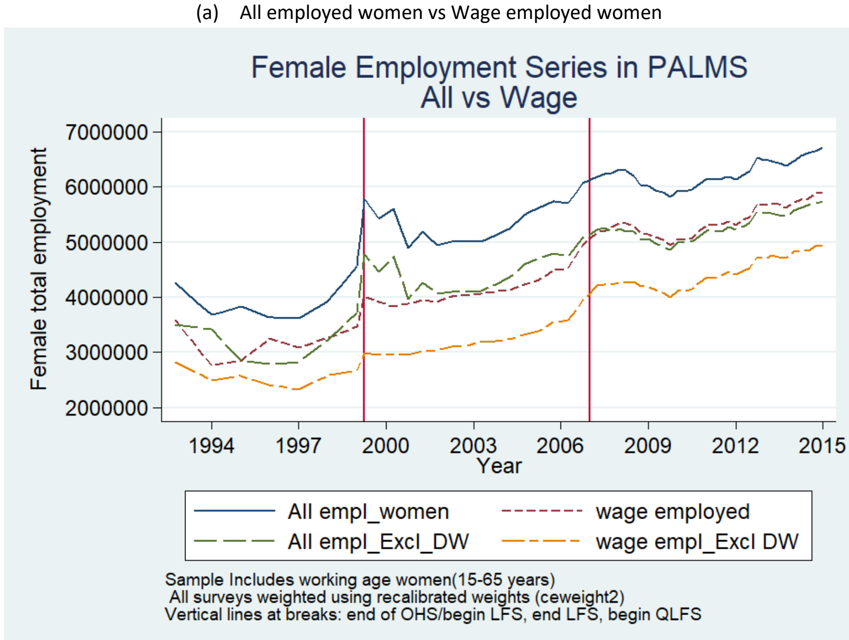
2.1.1 Trends in Female Employment in the PALMS Data Set

Figure 4 illustrates trends in female employment in the South African labour market from 1993 to 2015. There has been a continuous increase in female employment between 1994 and 2015. The figure compares trends in employment among all women with trends in employment among wage employed women (with and without domestic workers). Of interest to this analysis is the effect on these two trends when domestic workers are excluded. From the figure we see that domestic workers make up a significant proportion of women in employment. This can be seen by the shift in total employment figures when we exclude domestic workers.

The figure also shows a break in the data in the year 2000 and another one in 2008. These two periods also represent the changeover periods of survey instruments from the October Household Surveys (OHSs) to the Labour Force Surveys (LFSs) and from the LFSs to the Quarterly Labour Force Surveys (QLFSs). The figure confirms results from the literature that the much documented "feminisation" of the labour market in the early 2000s was partly as a result of the change in the survey instrument by Statistics South Africa (Stats SA) (Yu 2007, 2009; Branson & Wittenberg 2007; Posel & Casale 2001).

Figure 4a and 4b are similar except for the fact that we have excluded self-employed agricultural workers from figure 4b. Neyens & Wittenberg (2016) find that part of the big jump in 2000 was due to a large increase in the number of self-employed agricultural workers that resulted from the change of survey instrument by Stats SA. Excluding the self-employed agricultural workers leads to a smoother series although the jump is still visible.

(a) All employed women vs Wage employed women



(b) Excluding Self-employed agricultural workers

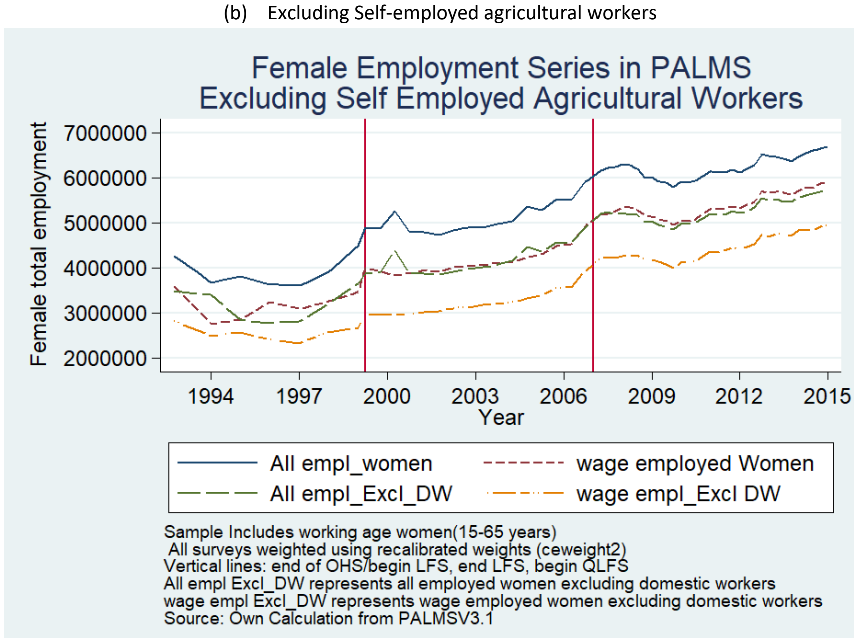
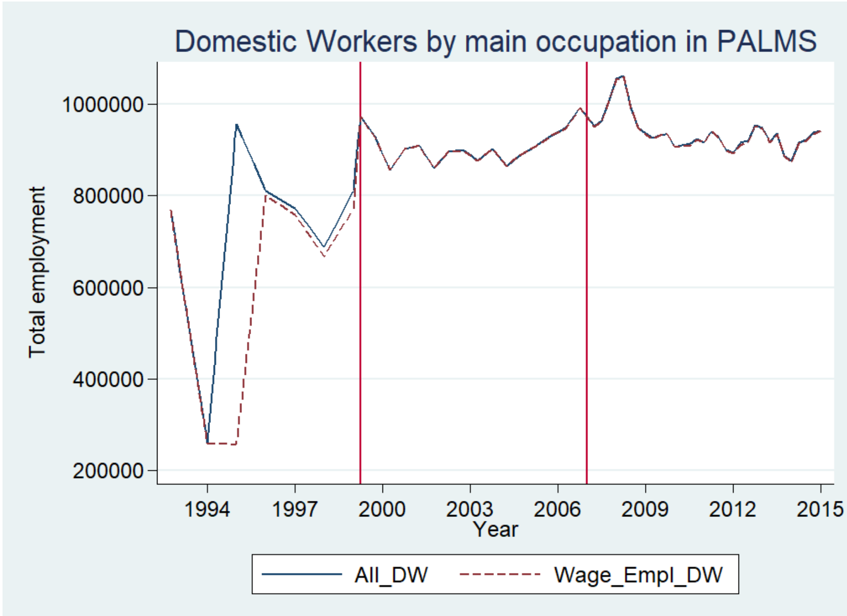


Figure 4: Employment Series of Women in PALMS

2.1.2 Trends in Domestic Services Employment in the PALMS Data Set

(a) Totals in the Domestic work Occupation



(b) Totals in the Domestic work Sector

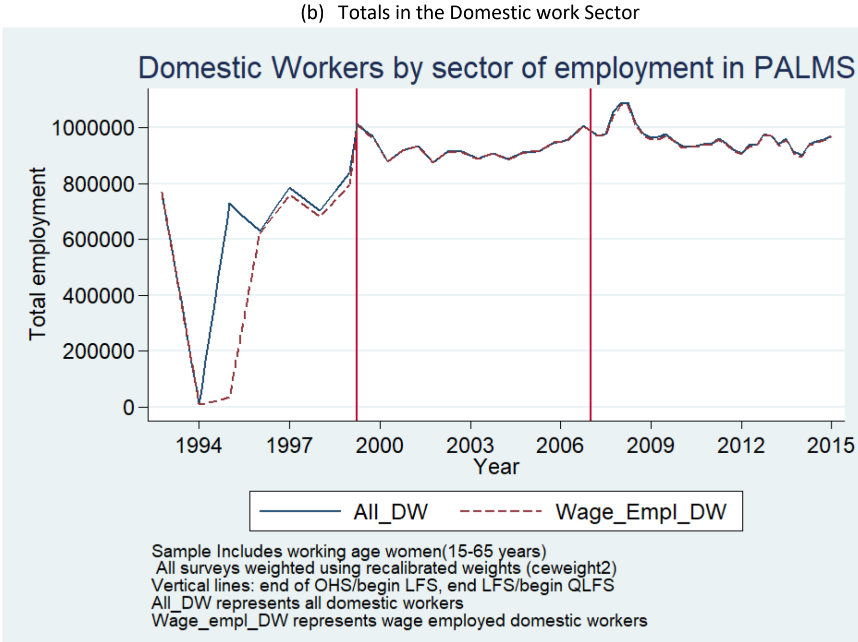


Figure 5: Employment Totals of Domestic Workers in PALMS

Source: Own Calculation from PALMSV3.1

Figure 5 details the number of women employed in domestic[11] work. The figure shows that apart from the OHSs (1994-1999), the total number of female domestic workers has remained between just over

[11] According to the Unemployment Insurance Contributions Act, 2002 (Act No. 4 of 2002), a domestic worker is “an employee who performs domestic work in the home of his employer, and includes a gardener, a driver or a person who takes care of any person in the home but does not include a farm worker”.

800,000 and 1 million. From the figure it is clear that how data on domestic workers was collected between the OHS 1994 and the OHS 1999 differed from the other data in the series.

Four issues arise from the trend exhibited by figure 5. One is that there are two variables in the national household surveys that can be used in the analysis of domestic work. That is, the main occupation variable and the industry variable. However, the totals one arrives at are going to vary depending on whether one uses the industry or the occupation variable especially between the OHS 1994 and the OHS 1999. This is because for some reason the OHS 1994, the OHS 1995 and the OHS 1996 contain a significant number of individuals reporting domestic work as the main occupation but not classified under the domestic services industry. For example, the OHS 1994 and the OHS 1995 contained about 808 and 820 observations respectively (about 489,829 and 497,104 weighted) reporting domestic work as the main occupation but not classified under the domestic services industry. Furthermore, as also noted by Budlender (2016, p.5 foot note 3) the industry variable seems to have workers that are clearly not "domestic workers" classified under this category for example professionals, technical and associate professionals. However, as illustrated in figure 5, this anomaly seems to decline after the OHSs, and the two variables give almost identical totals in the LFSs and the QLFSs.

The second issue is that there are fewer women with their main occupation classified as domestic work in all but the 1995 OHS. It therefore appears that we are missing some domestic workers between the OHS 1994 and the OHS 1999. A non-coverage problem that might have affected the domestic services sector has been highlighted in the literature (Kerr & Wittenberg 2015; Machemedze et al. 2014). Of interest to this chapter is that the most likely workers to be skipped were "live-in" domestic workers who were likely to be staying in back rooms on their employer’s compound (Machemedze et al. 2014). With the introduction of a master sample of enumeration areas in 1999, all households in multiple household dwellings were enumerated and this could explain the increase in domestic workers seen in figure 5 from 1999 onwards.

The third issue that arises from the trend in domestic workers employment is that there seems to be a change in how domestic workers were classified between the OHS 1994 and the OHS 1999. The *employerAll* variable is used in this analysis to identify whether an individual is self-employed, or wage employed. The gap in 1995 between wage employed domestic workers (Wage_Empl_DW) and all domestic workers (All_DW) is due to the fact that in 1995 most domestic workers are classified as self-employed. Interestingly, the questions[12] that were used to identify one's occupation remained similar in the OHSs (the OHS 1994-1999) therefore the inconsistency in the classification of domestic workers in 1994 and 1995 seems to have happened during the post-coding of information. The questions identifying the type of work one did were however more detailed in the LFSs and there was more probing[13] by field workers.

The fourth issue which has to do with the "feminisation" of the labour market, is that looking at figure 4b, the "wage employed women" series for all employed women mirrors the wage employment (wage_Empl_DW) series for domestic workers. In 2000, with the end of the OHS and the beginning of the LFS (illustrated by the first vertical line), there was a jump in the wage employment for all women of about 500,000 workers whereas there was a jump of about 200,000 domestic workers in the same period. Therefore, of the 500,000 wage employed women, 200,000 of them were domestic workers.

Studies on employment in the South African labour market using national household surveys have also reported increases in female employment in the early 2000s. Burger & Woolard (2005) using data from OHS 1995 and LFS 2002 find a 14.9 percent growth in female employment which was mostly in the informal sector and domestic services. Similarly, Casale et al. (2004) covering the period 1995-2003 concluded that it is likely that part of the two million jobs claimed to have been created between 1995 and 2003 are most likely the influence of better data collection by Stats SA. They point to the fact that questionnaires in the LFSs were designed to capture more informal work as there were criticisms that the informal sector in South Africa was under counted. They show that most of the jobs that were created in this period came from the low earning sectors of self-employment, informal sector and domestic work. Additionally, Yu (2007) in a very comprehensive analysis covering the period between 1995 and 2006 concludes that the jump in employment during this period was as a result of improved capturing of participation rather than a real shift in labour force participation.

[12] In the OHSs the question used to identify one's occupation is "What kind of work is/was …doing at his/her main job?" For example, in the OHS 1994 this was question 3.12 and in the OHS 1995 and 1997 this was question 3.15. To identify own account workers and wage employed workers, the question asked in the OHSs is "Does/did…work for him/herself (formal/informal) or does/did he/she work for someone else?"

[13] In the LFSs, to identify informal work and own account workers there is a follow up question "In…'s main work was he/she… 1=Working for someone else for pay? 2=Working for one or more private households as a domestic employee, gardener, security guard? 3=Working on his/her own or on a small household farm/plot or collecting natural products from the forest or sea? 4=Working on his/her own with a partner in any type of business including commercial farms? 5=Helping without pay in a household business?"

2.1.3 What Happened to the Gender Wage Gap in 1994 and 1995?

The lack of the gender wage gap exhibited in figure 3 is peculiar. Indeed, compared to 1993, it seems like the gender wage gap suddenly reversed in 1994 and 1995 and just as quickly returned in the subsequent years. Although no study has tried to investigate the source of this anomaly, several authors have flagged it in the literature. Heap (2008) in her analysis of earnings inequality for the period between 1995 to 2006, dropped the year 1995 from her analysis citing that the year was anomalous as she found no gender wage gap in this year. Hinks (2002) who focused on the OHS 1995, found that African women enjoyed a 10% wage advantage against men in 1995, a result he attributes to higher productivity for female workers in terms of them having, on average, higher education than African men. Hinks, however, cautions against the interpretation of this result as domestic workers were under-represented in the 1995 sample (Hinks 2002, p.2047). Muller (2009) dropped the OHS 1995 from the analysis after finding a negative wage gap in this year. Muller citing Hinks (2002) attributed the negative wage gap to the under sampling of domestic workers in this year (Muller 2009). She notes that inclusion of the OHS 1995 would have given the impression of a worsening wage gap over time.

In this section we give a preview of the analysis of the gender wage gap which is carried out in detail in chapter three. Table 1 shows results from an Oaxaca decomposition[14] for selected waves using the PALMS dataset before adjustment. The overall wage gap variable gives the average wage differential between men and women whereas the explained component is the mean increase in women's wages if they had the same characteristics as men and the unexplained component is the part of the gap that cannot be explained by differences in characteristics (it is the difference in returns to observable characteristics).

The table shows a persistent gender wage gap between 1993 and 1999 except for 1994 and 1995. In 1994 and 1995, there seems to be a negative (-0.0646 and -0.0796 respectively) gender wage gap in favour of women.

The purpose of this section is to give a complete picture of the problem with the OHS 1994 and the OHS 1995 and to show why re-examination of the gender wage gap in chapter three is important. These Oaxaca decomposition results show that not only was the raw gender wage gap reported in 1994 and

[14] We discuss this methodology in detail in section 3.3.2.2 of Chapter three

1995 inconsistent with other years, the unexplained and the explained gaps reported in these years were also inconsistent.

Table 1: Oaxaca decomposition results in PALMS before data adjustment

Variables/Wave	PSLSD 93	OHS 1994	OHS 1995	OHS 1997	OHS 1998	OHS 1999
Male log wage	2.418***	1.986***	2.224***	2.259***	2.188***	2.140***
	(0.0230)	(0.0154)	(0.0112)	(0.0149)	(0.0214)	(0.0196)
Female log wage	2.076***	2.050***	2.304***	2.013***	1.962***	1.871***
	(0.0260)	(0.0192)	(0.0136)	(0.0179)	(0.0281)	(0.0246)
Overall gap	0.343***	**-0.0646*** **	**-0.0796*** **	0.246***	0.226***	0.268***
	(0.0347)	(0.0246)	(0.0176)	(0.0233)	(0.0353)	(0.0314)
Explained gap	-0.0433*	**-0.194*** **	**-0.245*** **	-0.105***	-0.0655**	-0.0487*
	(0.0259)	(0.0181)	(0.0138)	(0.0184)	(0.0281)	(0.0258)
Unexplained gap	0.386***	**0.129*** **	**0.165*** **	0.351***	0.291***	0.317***
	(0.0259)	(0.0202)	(0.0135)	(0.0166)	(0.0288)	(0.0236)
Observations	6,323	12,161	19,953	15,830	8,716	10,907

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Note: Dependent variable: Log hourly wage, explanatory variables: potential experience, marital status, race, union, province, education
Source: Own calculation from PALMSV3.1

In this chapter, we propose two explanations for the missing gender wage gap in 1994 and 1995, the non-coverage problem highlighted in the literature in the early OHSs (Kerr & Wittenberg 2015; Machemedze et al. 2014) and that domestic workers were classified differently in 1994 and 1995. Addressing the coverage issues is beyond the scope of this analysis however, for a clear picture of how coverage may have affected the total number of domestic workers in the early OHSs and the difficulty in trying to resolve this issue, we provide a small discussion in section 2.2.2. The main contribution in this chapter, however, is addressing the classification of domestic workers in 1994 which resulted in a missing data problem. We apply both stochastic and multiple imputation techniques to solve the missing data issue and the resulting data reveals a substantial raw gender wage gap in 1994 and 1995. We discuss this in detail in section 2.3.

We then use the adjusted data to examine changes in the domestic services sector in the post-apartheid period. We find that total employment figures of domestic workers have remained stable over the years and that earnings in the domestic services increased after the implementation of minimum wage legislation. However, domestic workers remain the lowest paid workers. Another important finding is that even though there is no proof of a decline in the employment totals of domestic workers, domestic work seems to be shifting towards a part time set up which we link to the minimum wage legislation and domestic worker protection laws.

The rest of this chapter is organized as follows. In section 2.2 we review previous studies on domestic work in the South African labour market. We then discuss the data and methods of handling missing data in section 2.3. In section 2.4 we analyse changes in the domestic work sector and discuss the results and present the conclusions in section 2.5.

2.2 LITERATURE REVIEW

2.2.1 Previous Empirical Research on Domestic Work in South Africa

Historically, domestic work just like care work, was not considered productive work (Gaitskell et al. 1983; Cock 1980). This meant that especially in a patriarchal society such as South Africa, domestic workers existed outside labour market legislation. Cock (1980) details how domestic workers mostly black women worked under oppressive conditions. In her widely cited work *Maids and Madams: A Study in the Politics of Exploitation*, domestic work under apartheid is described as exploitative in terms of the low pay, long working hours, and instant dismissals. This situation was exacerbated by influx control laws that prevented free movement of black people in designated "white" neighbourhoods. This meant that a domestic worker was tied to one employer and changing jobs was extremely difficult.

The official recognition of domestic workers as workers came in 1993 when the Basic Conditions of Employment Act (BCEA) of 1983 which provides protection with respect to conditions of work (working hours, leave days, dismissal) was amended to include domestic workers (Budlender 2016). In 1995, the Labour Relations Act (LRA), which provides for organisational rights and mechanisms for dispute resolution, was extended to cover domestic workers. In terms of organisation, domestic workers still lag behind because the nature of their employment makes mobilisation difficult and hence unionisation in the domestic services sector continues to be minimal.

The inclusion of domestic workers in the group of workers covered by the BCEA paved the way for the establishment of sectoral determination number 7 that set a minimum wage for domestic workers. The minimum hourly wage for domestic workers was effected in November 2002. Domestic workers were also included in the scope of the Unemployment Insurance Act through the Unemployment Insurance Act 63 of 2001 and the Unemployment Insurance Act 4 of 2002 (Budlender 2010).

Although it seems, in terms of legal protection that domestic workers have made some progress in the post-apartheid labour market, studies find that the minimum wages set are still very low (Budlender 2010) and that domestic workers are still the lowest paid group in South Africa (Hertz 2005). This is exacerbated by the fact that enforcement in this sector is very difficult due to the nature[15] of domestic work (Hertz 2005; Dinkelman & Ranchhod 2010; Bhorat et al. 2012).

Studies on the effect of the minimum wage legislation are in agreement that the minimum wage legislation led to a rise in wages for domestic workers (Hertz 2005; Dinkelman & Ranchhod 2010; Bhorat et al. 2013). However, results on the effect of the minimum wage legislation on hours of work and total employment, differ. Regarding hours of work, Dinkelman & Ranchhod (2010) find no evidence of a decline in hours but Hertz (2005) and Bhorat et al. (2013) report a decline in working hours. The study by Bhorat et al. (2013) shows however, that the rise in wages had a bigger impact than the decline in hours leaving the domestic workers better off in terms of total earnings. For total employment, Dinkelman & Ranchhod (2010) and Bhorat et al. (2013) find no clear evidence of a decline in employment. Dinkelman & Ranchhod (2010) attribute the lack of a decline in employment to poor enforcement of the legislation.

Hertz (2005) however, using 7 waves of the LFS data (LFS 2001-LFS 2004), reports a decline in employment between 2001 and 2004 which he attributes to the minimum wage legislation. The difference in results could be because of different samples used by these studies. While all three studies used data from the LFSs, Dinkelman & Ranchhod (2010) only focused on African and Coloured domestic workers in urban areas.

2.2.2 Coverage issues in PALMS: Live-in Domestic Workers

Live-in domestic workers were very common during the apartheid era as domestic workers who were mostly black South Africans, left their rural homelands that were far from the urban areas to search for

[15] Domestic work takes place in private homes which are difficult to access because authorisation by the owner is required. Additionally, because of the imbalance of power between the domestic worker and the employer, the worker may not wish to cause any trouble for the employer for fear of losing their employment (Lund & Budlender 2009).

employment in white suburbs. It was difficult to find accommodation in the "white" areas due to pass laws at the time and therefore they lived in small back rooms in their employer’s compound (Cock 1980). Additionally, even though the pass laws were abolished in 1986, Budlender (2016) writes that this arrangement of domestic workers living in their employer’s compound continued to be convenient as accommodation in the urban areas is expensive.

In a recent paper on sampling and fieldwork methodology in the OHSs, Kerr & Wittenberg (2015) found that there were errors in sampling and specifically there was an under sampling of small households due to fieldwork practices in the early OHSs (1994-1998). The authors detail that before the OHS 1999, fieldworkers were responsible for listing all dwellings in a particular enumeration area and then for drawing a random sample of 10 households by probability proportional to size. This means that in the period prior to 1999, there is a high chance that small households never made it into the sample and there does not seem to be any weighting correction for this design (Kerr & Wittenberg 2015). The first master sample of enumeration areas was introduced in 1999 based on the 1996 census. Therefore 1999 was the first year when all households at a dwelling that had more than one household per dwelling were sampled (Kerr & Wittenberg 2015).

Below we show the trend of live-in domestic workers from 1993 to 2015. Figure 6 shows a clear break in the data between 1994 and 1999. Since there is no variable in our data that identifies live-in domestic workers, we try to identify live-in domestic workers by first creating a variable that identified an enumeration area as predominantly white if the proportion of white people living in the area was 50% and above. We then classified all domestic workers that reported living in predominantly white areas as live-in domestic workers. We are aware that this figure could be biased downwards since there has been an emergence of black employers post-apartheid. However, according to the Time Use Surveys of 2000 and 2010, white South Africans still employ the largest proportion of domestic workers (Budlender 2016).

Figure 6 shows the total count of live-in domestic workers in PALMS. Using 1993 as a benchmark, the figure shows very few live-in domestic workers between OHS 1994 and OHS 1999. We believe that this is partly as a result of missing domestic workers due to coverage issues mentioned above. The "Livein_Dom_C4" series illustrates the up weighting of the totals of live-in domestic workers using recalibrated[16] weights (Machemedze et al. 2014) to adjust for under sampling in these early OHSs. The figure shows that the weights increase the totals but there is still an under-sampling problem.

Looking at the upper bound (95% confidence interval) of the number of live-in domestic workers weighted by ceweight4, the total number of live-in domestic workers in 1994 is still too far from the figure of live-in domestic workers in 1993 however, the upper bound is closer to the figure of live-in domestic workers in 1999. What this tells us is that the ceweight4 weights are actually helpful in up weighting the under sampled households even though from the totals in 1993 we could still be missing some live-in domestic workers between 1994 and 1998.

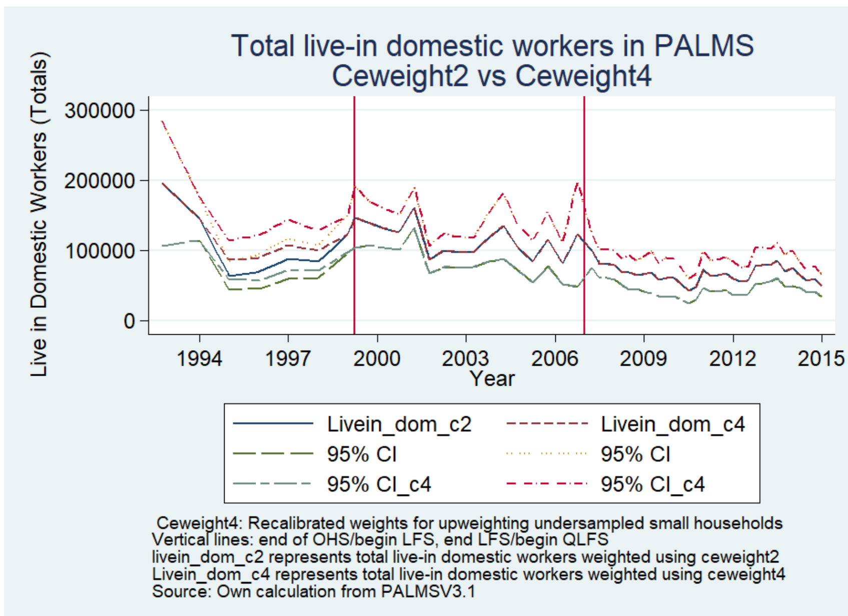


Figure 6: Weighted totals of live in Domestic Workers in PALMS

There has been a continued decline of the number of live-in domestic workers over the years. This trend can be attributed partly[17] to new legislation regarding working conditions for domestic workers. The domestic worker sectoral determinations of 2002 and 2011 outline new working conditions including acceptable working hours, minimum wage, leave and overtime payment (Budlender 2016). It is harder to draw the line in terms of normal working hours and overtime for live-in domestic workers and

[16] To upweight the small households in the earlier OHSs Machemedze et al. (2014) recalibrated the person weights released with the household surveys using demography, constant weights within households and total number of one person, two person and three-person households as constraints. They find that the reweighting produces big increases in the estimated number of domestic workers and mine workers in some of the earlier OHSs but also acknowledge that the missing data problem persists after the calibration exercise (see Machemedze et al. 2014, p.10,11).

[17] The decline of live-in domestic workers could also be due to the availability of alternative accommodation particularly in the urban areas for black women in the post-apartheid period and the conversion of domestic quarters.

therefore employers might therefore have opted for live-out domestic workers. Studies have found no real proof of job losses (Dinkelman & Ranchhod 2010; Bhorat et al. 2013) since the new legislation was effected but it is possible that employers opted for part-time or live-out workers over this period. Posel & Muller (2008) do report that over 50% of all part-time wage employed women are domestic workers.

2.3 DATA AND METHODS

2.3.1 Classification of Domestic Workers as "Self-employed" in 1994 and 1995.

For this analysis we utilise the Post-Apartheid Labour Market Series (PALMS) 1993-2015 (Kerr et al. 2016) discussed in section 1.3. We focus on all employed individuals between the ages of 15 and 65. For the imputation we focus on the PSLSD 1993, the OHS 1994 and the OHS 1995.

In 1994, we suspect that most domestic workers were classified as self-employed elementary workers while in 1995, domestic workers (or most of them) were categorized as own-account (self-employed) workers. After 1995, domestic workers were categorized as wage employees (see table 2).

Table 2 shows figures of total employment for domestic workers, services sector workers and elementary occupations workers by industry of employment and main occupation.

Table 2: Weighted Totals of Wage and Self-employment in PALMS

		Industry					
Services	**Employment type**	1994	1995	1996	1997	1998	1999
	Wage	1,100,595	1,075,907	1,024,397	836,139	924,648	912,994
	Self	**669,548**	37,825	21,830	29,898	40,033	70,682
Domestic Workers	Wage	**8,110**	**34,365**	619,479	758,708	681,487	796,465
	Self	**217**	**690,158**	1,085	11,743	348	424
		Main Occupation					
Elementary Occupations	wage	398,733	457,923	310,165	406,368	583,684	611,935
	self	**731,701**	67,591	47,133	107,935	110,248	209,586
Domestic Workers	Wage	**259,724**	**257,297**	800,265	758,708	667,189	771,077
	Self	**325**	**700,215**	0	712	348	424

Source: Own calculation from PALMS V3.1

Looking at the industry classification in table 2, we can see that the number of wage-employed domestic workers in 1994 (8,110) is far lower than other years, for example, 1996 (619, 479) or 1999 (796,465).

This corresponds to a very high number of services workers classified under self-employment (669,548). This figure is peculiarly high compared to other years. Even though we know that this deficit could be due to methodological and measurement error, the simplest explanation is that among the 669,548 self-employed service workers lie many domestic workers.

The similar classification difference is evident under the main occupation classification but in this case, domestic workers seem to be classified as self-employed elementary workers. For the OHS 1995, the number of wage-employed domestic workers is once again low but in this case they appear to have been classified as self-employed. That is, the number of self-employed domestic workers in 1995 is peculiarly high (700,215) compared to the other years which have fewer than 800 each.

Solving the classification inconsistency in the OHS 1995 is relatively simple as all we need to do is transfer the domestic workers classified under self-employment to wage employment (since they were still classified as domestic workers only under self-employment). The OHS 1994 is however slightly more complicated. The fact that domestic workers were classified under self-employed elementary or service sector workers makes it harder to transfer these domestic workers to the wage employment domestic work category as it is difficult to identify them.

To proceed, we treat this as a missing data problem where given the sampling distribution of wage-employed and self-employed domestic workers, elementary and service workers in 1993, 1994 and 1995, we assume that the domestic work variable contains missing values. We then apply both stochastic and multiple imputation techniques to impute for the missing values.

In the next section we briefly discuss the imputation methods used in this chapter to handle the missing data problem and then discuss the results from the imputations.

2.3.2 Handling Missing Data

2.3.2.1 Stochastic Imputation

There are several ways of handling missing data, for example, one can carry out a mean imputation which entails replacing the missing values for a variable with the mean of that variable calculated from the cases where values are not missing for that variable (Allison 2001). Although this method is used by researchers, the pitfall is that a mean imputation is likely to distort the variances and covariances or change the relationships between variables (Allison 2001). Alternatively, one could carry out a conditional mean imputation where predicted values from a regression model are used to replace the missing values. This however means that all predicted values fall directly on the regression line resulting in artificially high precision (Allison 2001; Wittenberg 2016b).

An exception is single stochastic imputation whereby it is possible to incorporate the lost variability into the imputation (Wittenberg 2016b). Unlike conditional mean imputation, single stochastic imputation adds a randomly drawn residual term from some distribution (normal, uniform or log-normal distribution) (see equation (1)) to imputed values from the regression imputation therefore adding back some of the lost variability.

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}x_i + e_i \tag{1}$$

The advantage of this method is that this can be done even when the actual data is sparse or non-existent. Additionally, since the distribution of residuals is based on residual variance from the regression model, single stochastic imputation restores some lost variability and produces unbiased coefficient estimates under the missing at random assumption (MAR). The drawback to this method however is that filled in data is still treated as real data and therefore the standard errors are artificially low (Allison 2001; White et al. 2011).

2.3.2.2 Multiple Imputation (MI)

As mentioned above, an advantage of single stochastic imputation over other imputation methods (unconditional mean imputation and single imputation) is that it deliberately adds variability to the data. However, the result is that this introduces two sources of variability. One is the within imputation variability resulting from the possibility of different samples being drawn from the same population and the other is between imputation variability resulting from the fact that even with the same sample, different random imputations are possible for missing data. Single stochastic imputation however is unable to capture the between imputation variability meaning that the resulting standard errors and confidence intervals will be too small (White et al. 2011).

Multiple Imputation therefore aims to correctly reproduce the variation and associations among the variables that would have been present in the full dataset by estimating both the variability in the population and the variability induced by imputation avoiding bias from missing data and resulting in valid statistical inference (Rubin 1996). The idea behind MI is that the single stochastic imputation is carried out several times and then each realized imputed data is analysed individually but identically to obtain a set of parameter estimates. The estimates are then combined using Rubin's combining rules to obtain overall estimates, variances and confidence intervals. Let θ be a set of parameters of interest given the data. The multiple estimates from each realisation of data($\hat{\theta}_1, \hat{\theta}_2, ... \hat{\theta}_m$) are averaged to the single $\hat{\theta}$ according to Rubin's rule as follows:

$$\hat{\theta} = \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j \tag{2}$$

where m is the number of imputations (in our case 20). The total variance of $\hat{\theta}$ is

$$var(\hat{\theta}) = W + (1 + \frac{1}{m})B \tag{3}$$

a combination of within imputation variance $W = (\frac{1}{m})\sum_{j=1}^{m} W_j$ and the between imputation variance $B = (\frac{1}{(m-1)})\sum_{j=1}^{m}(\hat{\theta}_j - \hat{\theta})^2$ (White et al.2011).

2.3.2.3 Imputation Model

To make adjustments to the data, we first reclassified all domestic workers in all the waves as wage employed. The reasoning here is that most, if not all, domestic workers provide services to someone else for a wage be it permanent, casual or part-time. This corrects for the figures in 1995 but not 1994 as these domestic workers were classified under elementary occupations.

For OHS 1994, we carried out a simple stochastic imputation in order to identify the domestic workers classified under elementary occupations. The procedure involved first calculating the probability of being a domestic worker using a pooled sample of elementary workers and domestic workers from the PSLSD 1993 and the OHS 1995 data. We picked the PSLSD 1993 and the OHS 1995 because they are the nearest distributions to the OHS 1994. For this we used a probit model

$$Pr\ (DW\ =\ 1|X)\ =\ \Phi(X\beta) \tag{4}$$

where DW is a binary dummy variable indicating that an individual is a domestic worker; Φ is the standard normal cumulative distribution function; X is a vector of covariates that includes an education categorical variable, a location dummy for urban and rural, a province categorical variable, age, age squared and gender. β represents the estimated coefficients.

We report results for two probit models. For the first probit model we use the main occupation variable and for the second probit model we use the industry of employment variable to construct the dummy variable for domestic worker.

We constructed the dummy variable for domestic worker (DW) from those who reported their main occupation to be domestic work ($jobocccode == 10$) for the first model and those who reported being in the domestic services sector ($jobindcode == 10$) for the second model. These individuals were given a value of 1 and all other individuals were given a value of 0. Only respondents that reported being employed in elementary or domestic work occupations in PSLSD 1993 and OHS 1995 and are aged between 15 and 65 years were used in estimating the first probit model and only those that reported being in the services and domestic services industries were used for estimating the second probit model.

2.3.3 Imputation Results

2.3.3.1 The Probability of Being a Domestic Worker in the PSLSD 1993 and the OHS 1995

The results from the imputation models are presented in table 3. The probit results for the first model (main occupation) reveal that females are 41.8% more likely to be domestic workers than men and respondents from all provinces are more likely to be domestic workers than their Western Cape counter parts. Relative to individuals in the rural areas individuals in the urban areas are 9.74% more likely to be domestic workers. The results also show that relative to individuals with no education individuals with some primary education are 10.8% less likely to be domestic workers and those with complete secondary education are 33.8% less likely to be domestic workers. Figure 7 shows that the effect of female differs with age. A 20-year-old female is 40% more likely to be a domestic worker whereas a 60 year old female is 44% more likely to be a domestic worker. The effect of urban however does not seem to vary with age.

The results from the second model ($Industry$) show that relative to individuals in the rural areas, urban dwellers are 9.62% less likely to work in the domestic services sector. Also, relative to having no education, completing secondary education reduces the probability of being in domestic work by 55.8% and attaining tertiary education reduces the probability of being in domestic work by 59.4%. The results also show that, as expected, women are 22.1% more likely to be in the domestic services sector than men.

Table 3: Imputation models: Probability of being a DW in the PSLSD1993 and the OHS1995

	Occupation		Industry	
Variables	Coefficients (S.E)	dy/dx (S.E)	Coefficients (S.E)	dy/dx (S.E)
urban	0.400***(0.0428)	0.0974***(0.0102)	-0.503***(0.0540)	-0.0962***(0.0102)
Some Primary	0.108**(0.0529)	0.0264**(0.0129)	-0.000205(0.0730)	-6.45e-05(0.0230)
Inco. Secondary	-0.00801(0.0586)	-0.00194(0.0142)	-0.865***(0.0741)	-0.277***(0.0232)
Complete SEC	-0.338***(0.108)	-0.0785***(0.0243)	-2.344***(0.106)	-0.558***(0.0219)
Some Tertiary	-0.622**(0.293)	-0.138**(0.0582)	-3.223***(0.172)	-0.594***(0.0210)
Degree_Higher			-3.088***(0.324)	-0.591***(0.0220)
Eastern Cape	0.477***(0.0809)	0.109***(0.0190)	0.194**(0.0892)	0.0373**(0.0172)
Northern Cape	0.317***(0.101)	0.0710***(0.0232)	0.184(0.181)	0.0353(0.0350)
Free State	0.792***(0.0656)	0.189***(0.0149)	0.501***(0.0883)	0.0976***(0.0172)
KZN	0.479***(0.0677)	0.110***(0.0154)	0.0680(0.0833)	0.0130(0.0159)
North West	0.737***(0.0741)	0.174***(0.0171)	-0.0304(0.104)	-0.00575(0.0196)
Gauteng	0.856***(0.0699)	0.205***(0.0169)	0.253***(0.0794)	0.0488***(0.0153)
Mpumalanga	0.818***(0.0779)	0.195***(0.0187)	0.395***(0.103)	0.0768***(0.0203)
Limpopo	0.395***(0.102)	0.0894***(0.0239)	-0.332***(0.111)	-0.0609***(0.0199)
female	1.718***(0.0421)	0.418***(0.00453)	1.156***(0.0554)	0.221***(0.00825)
age	0.0329***(0.0120)	0.00273***(0.000476)	-0.0647***(0.0144)	0.00287***(0.000425)
age2	-0.000291*(0.000149)		0.000643***(0.000180)	
Constant	-2.899***(0.239)		1.234***(0.294)	
Observations	12,858		12,279	
Pseudo R-sq	0.3283		0.4476	
Wald chi-Sq	1975.65		1496.72	
Prob>chi2	0.0000		0.0000	

Standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1

Source: Author's own calculation using PALMS V3.1

Notes: Marginal effects are estimated as the average of the individual marginal effects

Omitted groups: Rural, with no education, from the Western Cape and male.

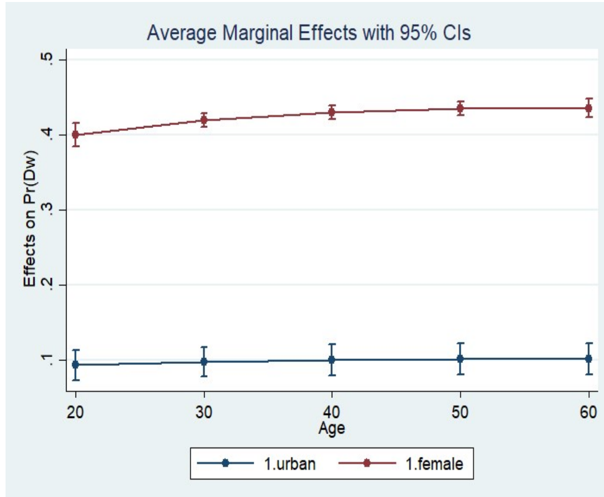


Figure 7: Average Marginal Effects by age
Source: Own calculation from PALMS V3.1

2.3.3.2 Single Stochastic Imputation Results

Using the probabilities estimated by the probit models above we carried out a one-time stochastic imputation of who might be a domestic worker in 1994. Our imputation procedure involved taking a random draw from the uniform distribution $U\ (0,1)$ for each individual in our sample. If the random number generated by a given draw was less than that individual's predicted probability of being a domestic worker, that individual was assigned domestic worker status. As a last step we reclassified all the self-employed domestic workers as wage employed domestic workers.

The new weighted totals are highlighted in bold in table 4. We estimate that there were about 951,823 domestic workers enumerated as elementary workers in OHS 1994 and about 704,245 domestic workers enumerated in the services industry. The adjusted employment and domestic worker totals are presented graphically in figures 8 and 9 respectively while the adjusted raw gender wage gap series is presented in figure 10.

The series still has many jumps telling us that the measurement issues are still not completely solved in this early period (OHS 1994-1999). However, the adjustment seems to have fixed the raw wage gap in 1994 and 1995. The new wage gap seems more plausible than the one in figure 3.

Table 4: Weighted Totals of Wage and Self-employment in PALMS after Adjustment

		Industry					
Services	Employment type	1994	1995	1996	1997	1998	1999
	Wage	923,056	1,079,702	1,024,994	836,139	924,648	912,994
	Self	**127,944**	34,542	21,830	29,898	40,033	70,682
Domestic Workers	Wage	**704,245**	**720,621**	623,515	782,308	698,497	810,380
	Self	**0**	**0**	0	0	0	0
		Main Occupation					
Elementary Occupations	wage	556,996	458,978	310,165	406,368	583,684	612,357
	self	**168,617**	65,086	46,741	107,935	108,744	206,519
Domestic Workers	Wage	**951,823**	**947,099**	803,803	769,914	683,197	782,663
	Self	**0**	**0**	0	0	0	0

Source: Own calculation from PALMS V3.1

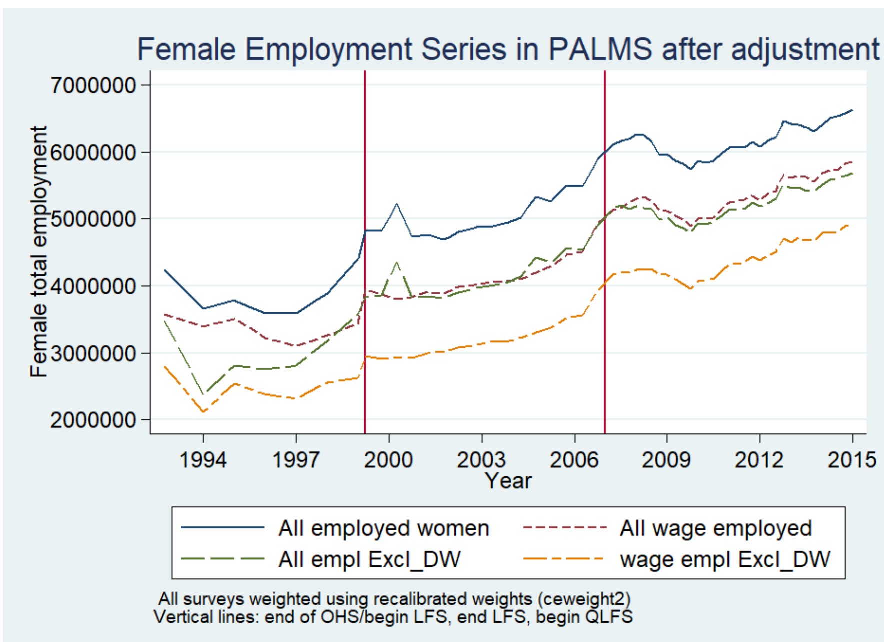


Figure 8: Total Female Employment in PALMS after adjustment

Source: Own calculation from PALMS V3.1

Notes: All empl Excl_DW represents all employed women excluding domestic workers.
wage empl Excl_DW represents wage employed women excluding domestic workers.

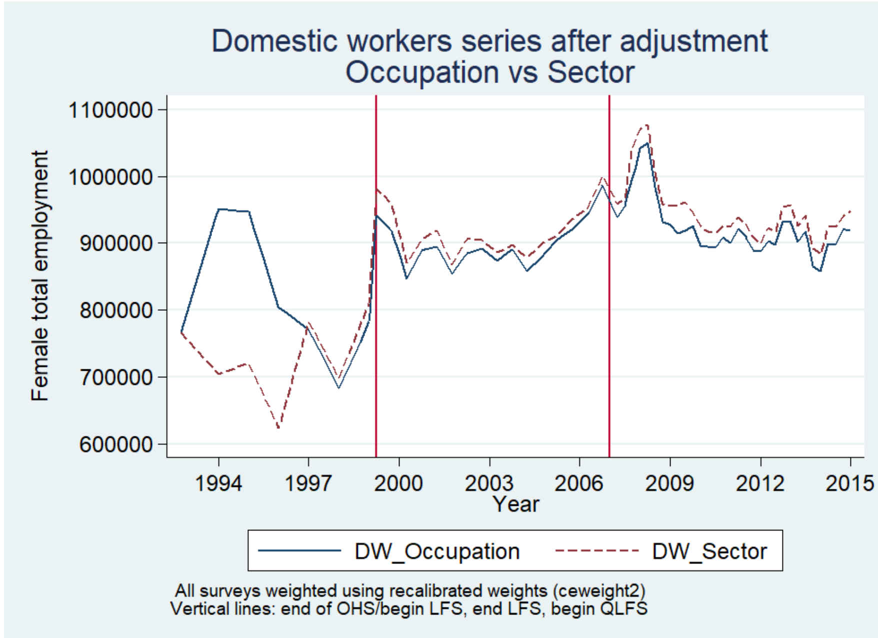


Figure 9: Totals for domestic workers in PALMS after adjustment

Source: Own calculation from PALMS V3.1

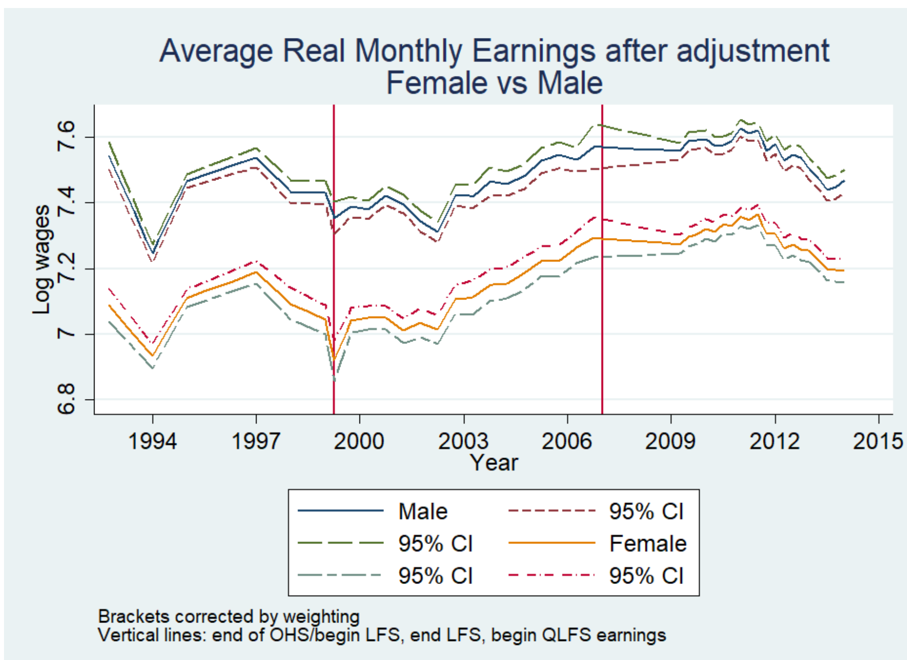


Figure 10: Raw wage gap after adjustment

Source: Own calculation from PALMS V3.1

2.3.3.3 Multiple Imputation Results

In table 5 we compare the estimates of the total number of domestic workers in 1994 from single stochastic imputation and those calculated using Multiple Imputation (MI). M1-M20 are the individual draws from the 20 draws of the multiple imputation. The total column gives the total number of domestic workers calculated from each realization of data. These are then averaged using the Rubin rules detailed in section 2.3.2.2 and the result is the *mi_estimate*.

Table 5: Total Female domestic workers in 1994 Single vs Multiple Imputation (20) draws

	Industry					**Occupation**				
Draws	Total	Std.Err	95% CI	95% CI	Obs.	Total	Std.Err	95% CI	95% CI	Obs.
M1	682,834	32,176	619,696	745,973	2,752	900,364	38,119	825,562	975,166	3,681
M2	706,732	33,232	641,520	771,943	2,883	879,483	37,606	805,688	953,277	3,615
M3	681,963	32,913	617,378	746,548	2,696	883,318	36,447	811,799	954,838	3,567
M4	704,460	32,088	641,494	767,427	2,943	854,960	36,850	782,649	927,271	3,474
M5	693,650	32,734	629,416	757,885	2,815	894,673	37,893	820,316	969,031	3,527
M6	703,681	32,198	640,498	766,864	2,924	867,427	36,910	794,999	939,856	3,519
M7	697,487	32,441	633,828	761,146	2,837	876,197	37,419	802,769	949,624	3,557
M8	693,928	32,919	629,330	758,525	2,828	863,744	37,913	789,346	938,141	3,500
M9	674,979	32,051	612,085	737,872	2,828	888,597	37,413	815,181	962,013	3,551
M10	705,572	32,093	642,595	768,548	2,954	876,928	39,817	798,794	955,061	3,457
M11	684,659	31,160	623,513	745,805	2,857	888,838	37,799	814,664	963,011	3,577
M12	679,643	32,704	615,467	743,818	2,803	862,529	36,755	790,405	934,653	3,610
M13	698,886	33,284	633,573	764,198	2,863	892,283	38,445	816,843	967,724	3,577
M14	680,311	32,242	617,042	743,579	2,801	880,598	37,871	806,284	954,912	3,559
M15	699,075	33,062	634,197	763,954	2,834	878,349	38,309	803,176	953,522	3,495
M16	695,469	33,337	630,051	760,887	2,824	868,385	36,650	796,467	940,302	3,515
M17	681,735	33,050	616,880	746,590	2,764	875,309	37,806	801,121	949,497	3,459
M18	693,939	31,743	631,650	756,228	2,880	883,239	37,267	810,109	956,369	3,522
M19	686,910	33,511	621,152	752,669	2,703	877,192	37,633	803,345	951,039	3,579
M20	697,789	32,766	633,492	762,087	2,816	865,336	36,785	793,154	937,519	3,527
Mi_esti	**692,185**	**34,102**	**625,226**	**759,144**	**2,752**	**877,887**	**39,477**	**800,370**	**955,405**	**3,681**
Single_est	**704,245**	**32,413**	**640,716**	**767,774**		**951,823**	**39,721**	**873,971**	**1,029,675**	

Source: Own calculation from PALMS V3.1
Note: M1-M20 are the individual draws from the 20 draws of the multiple imputation.
The Total column gives the total number of domestic workers calculated from each realization of data.
mi_est gives the average estimate calculated according to the Rubin Rules
Single_est is the estimate from the single stochastic imputation

The result from the multiple imputation shows that there were approximately 692,185 domestic workers in the domestic work sector while the figure from the single stochastic imputation is 704,245. Both these figures are within the confidence intervals of each method. As expected, the standard errors from the multiple imputation are bigger than the standard errors from the single stochastic imputation. Similarly, using multiple imputations we estimated 877,887 workers under the domestic work occupation compared to 951,823 domestic workers using single stochastic imputation. Both these estimates lie within the confidence intervals of each method and the standard errors from the multiple imputation are very similar to the standard errors from the single stochastic imputation. Given these results, the rest of the analysis is carried out using the OHS 1994 data from the stochastic imputation.

2.4 CHANGES IN THE DOMESTIC SERVICES SECTOR

From here on we utilise the resulting data after adjusting for the inconsistencies in the OHS 1994

2.4.1 Descriptive Statistics: Characteristics of Workers in the Domestic Services Sector

Table 6 presents summary statistics from PALMS data and shows that domestic work in South Africa is still mostly carried out by black women with 5 to 8 years of education. Over 90% of the female domestic workers are black and the remaining 10% are almost usually coloured. This trend has not changed much over the years. The level of education of domestic workers seems to have increased slowly over the years from just about 5 years of education in the 1990s to about 8 years of education in the 2000s. The average age of domestic workers has also increased gradually over the years. The average age of a domestic worker in the 1990s was 38 years and currently the average age of a domestic worker is about 42 years, meaning that young people are less likely to be in domestic work. This could be linked to the fact that more young people are pursuing education beyond primary school thus increasing their chances of employment in other sectors that are more skill intensive.

The table also shows that in the 1990s domestic workers were more likely to be married. However, this trend seems to be declining and the number of domestic workers who have never been married is on the rise. This is consistent with literature on the rise of female headed households in South Africa and the rise of households with women as the only income earners (Casale 2004). Domestic workers also come from all provinces. The summary statistics show that Gauteng hosts the largest number of domestic workers followed by Kwazulu-Natal. This is because these are the most populous provinces and the fact that these provinces are mostly urban with Gauteng being home to both Johannesburg and Pretoria.

Table 6: Summary Statistics- Female Domestic Workers in PALMS after data adjustment

Variable	**1993**	**1994**	**1995**	**1999**	**2002B**	**2005B**	**2007B**	**2010D**	**2012D**	**2014D**
Mean yrseduc	5.23	6.19	5.24	5.81	6.11	6.83	6.92	7.62	7.9	8.05
Age	39.13	38.2	38.12	39.56	40.14	40.03	40.84	42.12	42.44	42.71
Log monthly wage	6.13	6.28	5.82	5.96	5.82	6.16	6.31	6.32	6.34	6.34
Hourly wage	4.42	5.27	3.28	5.35	2.87	3.68	4.57	5.2	5.51	5.4
Hours (weekly)	38.06	41.32	40.91	41.02	41.44	41.21	38.42	34.45	33.9	35.64
Race										
African	0.93	0.83	0.88	0.88	0.9	0.92	0.91	0.93	0.91	0.92
Coloured	0.07	0.09	0.12	0.11	0.1	0.08	0.09	0.07	0.08	0.07
Indian	0.01	0	0	0	0	0	0	0	0	0
White	0	0.07	0	0	0	0	0	0	0.01	0
Marital Status										
Married	0.44	0.54	0.59	0.48	0.44	0.42	0.43	0.41	0.4	0.43
Widow	0.12	0.09	0.09	0.08	0.11	0.1	0.11	0.1	0.09	0.09
Divorced	0.24	0.06	0.06	0.09	0.09	0.07	0.07	0.05	0.05	0.04
Never_Married	0.2	0.31	0.26	0.35	0.37	0.42	0.4	0.44	0.46	0.44
Province										
Western Cape	0.08	0.08	0.11	0.12	0.12	0.12	0.11	0.1	0.11	0.1
Eastern Cape	0.11	0.11	0.12	0.13	0.14	0.12	0.1	0.1	0.09	0.1
Northern Cape	0.01	0.02	0.03	0.03	0.04	0.02	0.02	0.03	0.03	0.02
Free State	0.11	0.18	0.2	0.09	0.1	0.07	0.06	0.07	0.08	0.06
KwaZulu Natal	0.13	0.18	0.17	0.17	0.21	0.16	0.17	0.18	0.19	0.2
North West	0.11	0.09	0.09	0.08	0.08	0.09	0.07	0.06	0.05	0.08
Gauteng	0.3	0.21	0.16	0.24	0.18	0.28	0.31	0.29	0.28	0.26
Mpumalanga	0.1	0.07	0.08	0.08	0.07	0.08	0.08	0.08	0.08	0.08
Limpopo	0.05	0.05	0.06	0.07	0.07	0.06	0.08	0.09	0.09	0.09

Own calculation from PALMS v3.1

Figures 11 and 12 compare weekly working hours and monthly earnings of domestic workers to all wage employed women. Although the wage trend in the early OHSs is not very clear, the trend displayed from the early 2000s seems to be more consistent. Figure 11 shows that hours worked per week have been on the decline. However, the decline is much sharper for domestic workers. Hours worked for domestic workers dropped from an average of about 42 hours in 2000 to about 35 hours per week in 2015. Interestingly this drop took place between 2002 and 2009 and seems to have stabilised after that. Figure 12 on the other hand shows that monthly earnings for domestic workers have been on an upward trend since 2002. The rise in wages coincides with the introduction of minimum wage legislation for domestic workers and is well documented (Dinkelman & Ranchhod 2010; Hertz 2005). The decline in hours worked points to a shift towards part-time work for domestic workers which may explain why the total number of employees in the domestic services sector has remained stable over the years.

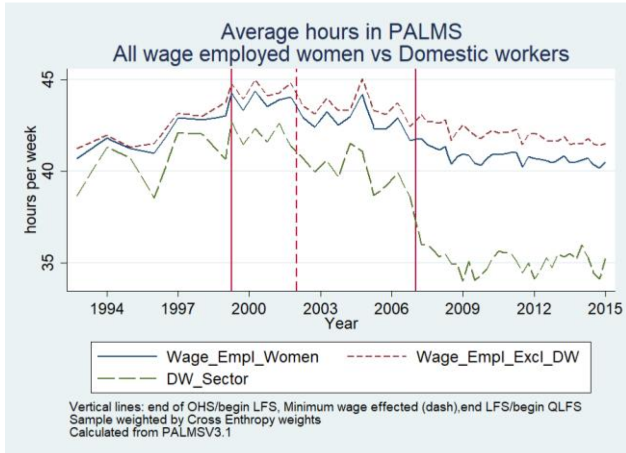


Figure 11: Average hours worked in the past week in PALM

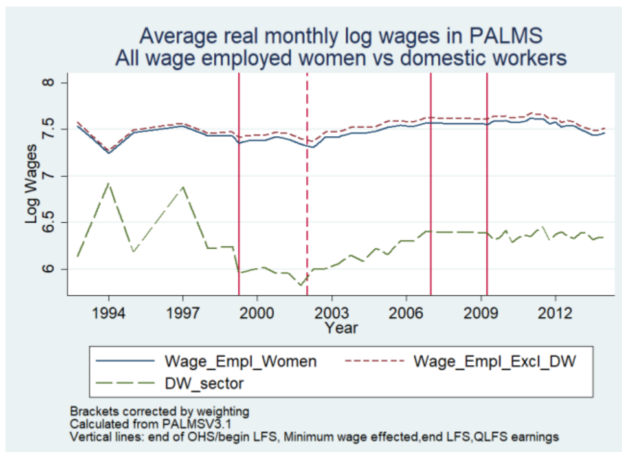


Figure 12: Mean Monthly Earnings for women in PALMS

2.4.2 Who is a Domestic Worker in South Africa?

Having looked at the measurement issues in PALMS and how domestic workers help explain employment trends, we investigate the determinants of being in the domestic services sector using a probit model. For this model the explanatory variables[18]include a location dummy($1 = urban, 0 = rural$), a married dummy($1 = married, 0 = divorced, widowed, never\ married$), a race dummy($1 = African, 0 = Coloured$), a 4 category education variable, a 9 province categorical variable, and a 3 category age variable as explanatory variables. We constructed the dummy variable for domestic worker (DW) from those who reported being in the domestic services sector($jobindcode == 10$). These individuals were given a value of 1 and all other individuals were given a value of 0. Only wage employed respondents

[18] The following categories were dropped from the analysis for lack of observations: the Asian/Indian and white race category and anyone with education higher than secondary education.

who are aged between 15 and 65 years were used in estimating the probit model. The results are presented in table 7.

The results show that being older and moving from rural to urban increases the probability of being in the domestic services sector for women in South Africa. They also show that having primary and incomplete secondary education as opposed to having complete secondary education and being black increases the probability of being in the domestic services sector. These results are consistent with the descriptive statistics that the South African domestic worker is most likely to be female, black, with low education and much older in age (in her late 30s or 40s). This tells the story that the domestic services sector has not changed much in the post-apartheid period.

2.5 DISCUSSION

This chapter sought to show that domestic workers are key to understanding employment trends for women in PALMS. We showed that the under sampling of live-in domestic workers can partly explain the low employment figures in the early OHSs and that the introduction of a master sample of enumeration areas that led to better capture of this group can be linked to the apparent "feminisation" of the labour market. This is in agreement with the literature that cites that most of the employment in the early 2000s came from the informal sector and domestic work (Burger & Woolard 2005; Casale et al. 2004).

In this analysis we have also illustrated the reasons why the 1994 and the 1995 data exhibit no raw gender wage gap. We have shown that this "missing" raw gender gap was due to the classification of domestic workers as self-employed workers in the OHS 1994 and the OHS 1995. We have suggested an adjustment and shown the impact of the adjustment on the employment trends and the raw gender wage gap, importantly, the adjustment brings back a plausible gender wage gap consistent with other years in the series. This shows that understanding how domestic work was captured in the beginning of the early OHSs is important in understanding inconsistencies in the household surveys and trends in employment.

After addressing these measurement issues using imputation methods, we then looked at the changes in the domestic sector in the post-apartheid period. We find that total figures in employment of domestic workers have remained stable over the years and conditions of employment for domestic workers have improved over the period 1993-2015 in terms of working hours and earnings. However, domestic workers remain the lowest paid workers. Another important finding is that even though there is no proof of a decline in the employment totals of domestic workers, domestic work seems to be shifting towards a predominantly part-time occupation linked to the minimum wage legislation and domestic worker protection laws.

The data problem bares on two issues. One is that missing domestic workers affects the aggregate employment and wage employment levels. The second issue is the issue of a non-existent gender wage gap. As we have shown in this chapter, data issues relating to missing domestic workers are not only important for the domestic work story which is interesting in and of itself but also affects statistics such as the gender wage gap. Most studies that used 1995 as their baseline survey ended up concluding that the wage gap rose in the period between 1995 and the 2000s. The implication is that the choice of baseline can determine substantial interpretation of results. In the next chapter we analyse the gender wage gap in detail using a combination of methods and update the results of the gender wage gap found in the literature.

Table 7: Probit results: Determinants of being a domestic worker

Table 7: Probit results: Determinants of being in the domestic services sector in PALMS

Wave	OHS1994	OHS1995	OHS1996	OHS1998	OHS1999	LFS00:2	LFS01:2	LFS02:2	LFS03:2	LFS04:1
Urban	-0.0423***	-0.0657***	-0.0340*	0.0304*	0.0413***	0.0568***	0.00828	0.0290*	0.0413***	0.0755***
	(0.0123)	(0.0119)	(0.0189)	(0.0181)	(0.0151)	(0.0157)	(0.0150)	(0.0155)	(0.0153)	(0.0191)
Married	-0.00739	0.00720	-0.0119	-0.0303*	-0.0219*	-0.00567	-0.0357***	-0.00180	-0.0281**	-0.0219
	(0.0119)	(0.0110)	(0.0157)	(0.0155)	(0.0130)	(0.0142)	(0.0131)	(0.0136)	(0.0134)	(0.0162)
ec	0.0336	0.0201	-0.0120	-0.000888	0.120***	0.0246	0.0380	0.0381	0.0299	0.0944***
	(0.0214)	(0.0226)	(0.0355)	(0.0341)	(0.0288)	(0.0313)	(0.0297)	(0.0291)	(0.0290)	(0.0359)
nc	0.189***	0.139***	0.0911***	0.0538*	0.134***	0.183***	0.140***	0.139***	0.109***	0.0752
	(0.0242)	(0.0275)	(0.0351)	(0.0316)	(0.0303)	(0.0309)	(0.0311)	(0.0305)	(0.0308)	(0.0623)
fs	0.182***	0.209***	0.0432	0.0317	0.0151	-0.0161	-0.00465	-0.0227	-0.0141	0.0111
	(0.0241)	(0.0232)	(0.0359)	(0.0344)	(0.0305)	(0.0327)	(0.0307)	(0.0305)	(0.0318)	(0.0371)
kzn	-0.0377*	0.0124	-0.0651*	-0.124***	-0.0316	-0.0128	-0.0238	-0.0438	-0.00379	0.0229
	(0.0227)	(0.0241)	(0.0368)	(0.0340)	(0.0301)	(0.0335)	(0.0293)	(0.0289)	(0.0292)	(0.0342)
nw	0.00397	0.0355	-0.0730*	-0.0415	-0.000149	0.0450	0.0154	-0.0457	0.00206	0.0435
	(0.0261)	(0.0268)	(0.0393)	(0.0345)	(0.0304)	(0.0333)	(0.0319)	(0.0313)	(0.0322)	(0.0366)
Gauteng	0.0128	-0.0184	0.0105	-0.0896***	0.0364	0.0409	0.0409	-0.0752***	-0.0288	0.00919
	(0.0235)	(0.0251)	(0.0324)	(0.0330)	(0.0280)	(0.0308)	(0.0286)	(0.0288)	(0.0292)	(0.0352)
Mpl	-0.0117	0.00917	-0.0724*	-0.0530	-0.00968	-0.0402	-0.0412	-0.0494	-0.0993***	-0.0495
	(0.0292)	(0.0267)	(0.0386)	(0.0352)	(0.0317)	(0.0355)	(0.0336)	(0.0332)	(0.0332)	(0.0382)
Limpopo	-0.0590*	-0.0806***	-0.205***	-0.143***	-0.0699**	-0.109***	-0.130***	-0.155***	-0.104***	-0.0606
	(0.0314)	(0.0312)	(0.0443)	(0.0392)	(0.0350)	(0.0366)	(0.0351)	(0.0331)	(0.0341)	(0.0404)
31-45 yrs	0.0182	0.0106	0.0150	0.0531***	0.0180	-0.00773	0.00946	0.00768	0.0519***	0.00765
	(0.0138)	(0.0131)	(0.0199)	(0.0191)	(0.0161)	(0.0179)	(0.0163)	(0.0172)	(0.0166)	(0.0209)
46-65 yrs	0.0407**	0.0415***	0.0235	0.0598***	0.0812***	0.0187	0.0324*	0.0275	0.0508***	0.0441*
	(0.0170)	(0.0156)	(0.0233)	(0.0220)	(0.0194)	(0.0202)	(0.0189)	(0.0198)	(0.0194)	(0.0229)
primary	-0.00176	-0.0366**	-0.0231	-0.0346	-0.0429**	-0.0435*	-0.0417*	-0.0377*	-0.0664***	-0.0114
	(0.0186)	(0.0163)	(0.0252)	(0.0220)	(0.0199)	(0.0224)	(0.0214)	(0.0217)	(0.0234)	(0.0297)
Inc_sec	-0.154***	-0.220***	-0.227***	-0.188***	-0.243***	-0.231***	-0.225***	-0.232***	-0.240***	-0.186***
	(0.0191)	(0.0169)	(0.0256)	(0.0224)	(0.0204)	(0.0229)	(0.0214)	(0.0221)	(0.0232)	(0.0298)
Matric	-0.423***	-0.467***	-0.510***	-0.451***	-0.493***	-0.481***	-0.479***	-0.462***	-0.477***	-0.427***
	(0.0295)	(0.0247)	(0.0341)	(0.0313)	(0.0251)	(0.0296)	(0.0252)	(0.0256)	(0.0257)	(0.0335)
African	0.138***	0.0993***	0.163***	0.206***	0.150***	0.226***	0.184***	0.236***	0.230***	0.201***
	(0.0175)	(0.0186)	(0.0286)	(0.0274)	(0.0241)	(0.0255)	(0.0236)	(0.0239)	(0.0246)	(0.0306)
Obs	8,299	7,902	3,728	4,399	6,353	6,991	6,589	6,124	6,031	6,101

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Notes: Omitted groups: Coloured, from the rural area, with no schooling, from western cape , aged between 15 and 30 years, and not married. Includes working age (15-65) years and wage employed.Marginal effects are estimated as the average of the individual marginal effects

3 DISTRIBUTIONAL CHANGES IN THE GENDER WAGE GAP

3.1 INTRODUCTION

Great strides have been made in enhancing gender equality over the past few decades both in the developed and the developing world. Anti-discrimination legislation and various "Girl child" initiatives including the global Millennium Development Goals [19](MDGs), have led to some gender gaps in human capital characteristics being closed such as education. Both international and local studies report increased labour force participation of women attributed to inter alia the closure of the gender gap in education (Ganguli et al. 2014), advances in medicine such as the introduction of contraceptive pills (Goldin & Katz 2002) and technological progress both in the labour market and in the home (computers and microwave ovens) (Blau & Kahn 2000; Petrongolo & Olivetti 2006).

In South Africa, Ntuli & Wittenberg (2013) report that increased labour force participation of black women is attributable to changes in social norms and changes in women's behavioural response towards the labour market rather than to changes in human capital characteristics. Better access to education has increased women's opportunity cost of being outside the labour market and therefore increased their probability of participation. Also, as Casale & Posel (2002) find, the proportion of women living with employed men has declined over time, a finding consistent with Casale (2004) who reports that the proportion of households dependent on women's earnings increased from 14.8 percent to 21 percent in the period between 1995 and 2001. This suggests that the traditional notion of a male breadwinner is weakening in the South African context. Casale & Posel (2002) also report that increased labour force participation of women can be attributed to the decline in marriage and fertility rates and an increase in education.

Deliberate anti-discrimination legislation by the newly elected democratic government in 1994 has also contributed to increased female labour force participation in South Africa (Burger & Jafta 2010). These policies include the *Labour Relations Act, number 66 of 1995* which governs how employers and employees interact in the work place, the *Basic Conditions of Employment Act, number 75 of 1997* which

[19] The MDGs were the world's time-bound and quantified targets for addressing extreme poverty in its many dimensions while promoting gender equality, education, and environmental sustainability.

regulates working conditions including working hours and allows the minister of labour to determine minimum wages for employees by sector, the *Employment Equity Act number 55 of 1998* which was aims to promote equal opportunity and fair treatment in employment through the elimination of unfair discrimination and the *Black Economic Empowerment Act, number 53 of 2003* whose objectives were to facilitate broad-based black economic transformation in order to enable meaningful participation of black people in the economy (Leibbrandt et al. 2010a). These policies have contributed to the rise in women's level of participation by increasing the returns to women's labour force characteristics (Ntuli & Wittenberg 2013).

According to the human capital theory which attributes differences in wages between men and women to differences in educational qualifications and labour market experience (Mincer 1974; Polachek 2006), the above developments imply that there should have been a convergence of earnings between men and women. However, the gender wage gap persists in many countries and a large portion of it cannot be explained by differences in human capital characteristics (Hirsch 2016). Polachek (2014) reports that despite the introduction of equal pay legislation in the United States of America in 1963, as at 2014, American women still earned 22% less than men. He also points out that in the United Kingdom and France women earned 21% and 17% less than men respectively in the same period (Polachek 2014) despite there being equal pay legislation in those countries. Goldin (2014) reports that even after controlling for differences in human capital characteristics a substantial gender wage gap remains. She attributes the persistent gender wage gap to differences in time demands in different industries.

In the previous chapter, we showed that issues in the classification of domestic workers led to a wage gap in favour of women in 1994 and 1995. This led to researchers reporting mixed results on the trend of the gender wage gap over time with some finding a rise in the gender wage gap between 1995 and 2006 and others reporting a drop depending on the choice of the base period. Additionally, it is not clear how issues such as earnings given in brackets and outliers were dealt with in the previous studies. These data quality issues that have come to light as more household surveys became available, necessitate the re-examination of the gender wage gap in South Africa.

This chapter examines the evolution of the gender wage gap over the period 1993-2015 using the Post-Apartheid Labour Market Series (PALMS) data corrected for domestic worker misclassification in 1994 and 1995 as described in chapter 2. We ask whether the unexplained gap which is associated with discrimination has declined over time. If it has been declining over time, what are some of the most important characteristics associated with the decline? In particular we assess whether affirmative action been binding and the implication on the gender wage gap across the wage distribution.

The results show that the mean gender wage gap narrowed significantly during the period studied, from 0.34 log points (about 40%) in 1993 to 0.15 log points (about 16%) in 2014. Distributional analysis shows that the median wage gap which is greater than the mean wage gap has stagnated over time staying between 35% and 23% with some fluctuations. There has been a substantial decline in the gender wage gap at the 10th percentile from 0.47 log points (about 60%) in 1997 to 0.18 log points in 2010 (about 19%). We attribute the narrowing of the gender gap at the bottom of the wage distribution to minimum wage legislation while the stagnation of the gender gap at the median suggests that anti-discrimination legislation has not been binding. Also, we find that education is an important factor in reducing the unexplained gap at the 10th percentile but not at the 90th percentile.

The rest of this study is organized as follows in section 3.2 we review economic theory that has been used to explain the gender wage gap and discuss previous empirical studies on the gender wage in the South African labour market. We then discuss the data and methods that will be used in this analysis in section 3.3. Section 3.4 outlines the methodological issues that may be encountered during this analysis. We present the results in section 3.5 and finally we discuss these results and present our conclusion in section 3.6.

3.2 LITERATURE REVIEW

3.2.1 Economic Theories Explaining the Gender Wage Gap

According to the human capital theory, wage differences between men and women arise from the fact that men and women have different educational qualifications and different years of experience (Mincer & Polacheck 1974; Altonji & Blank 1999). Generally, the view is that given the traditional division of labour, women spend less time in the labour market and therefore they accumulate less labour market experience (Polachek 2006). Similarly, because they anticipate spending less time in the labour market, they are more likely to invest less in formal education and on-the-job market training (Polachek 2006). This therefore means that women are more likely to be employed in jobs that are less skill intensive and which are more flexible time wise. These jobs in turn tend to pay less. The argument put forward therefore, is that if there are high returns to human capital characteristics regardless of gender, the lower the stock of human capital characteristics attained by women, the higher the wage gap.

Also used to explain the gender wage gap are the economic theories on discrimination which include the *Taste based discrimination theory* by Becker (1971) and *Statistical discrimination theory* whose advocates include Arrow (1973) and Coate & Loury (1993). According to Aigner & Cain (1977) discrimination occurs when equal productivity is not rewarded with equal pay.

According to Becker (1971) the wage gap in the labour market can be explained by the fact that employers have a 'taste' for discrimination. Becker's taste-based model assumes that employers, employees and customers are prejudiced and that they are willing to pay a price not to encounter a certain group in our case females (Altonji & Blank 1999). Assume there are two groups in the labour market, group A, men and group B, women. Employers maximize their utility function from employing members of their preference group in this case, men. Critics of this theory postulate that perfect competition should eliminate discrimination. This is because discrimination is costly and goes against the firms' objective to maximize profit. Assuming free entry of firms and constant returns to scale, non-discriminating firms will hire members of group B at a lower wage and make more profits thereby pushing discriminating firms out of the market. If there are enough firms to hire all members of group B, then there will be no wage gap. Employee taste-based discrimination is when male employees do not want to encounter female employees and need to be compensated to work with females whereas customer prejudice is when customers do not want to encounter workers of a certain group (females). Unlike the case of employer prejudice where competition can end discrimination, customer prejudice may propagate long-term discrimination and lead to segregated labour markets.

As discussed by Altonji & Blank (1999), statistical discrimination stems from the view that employers have limited information about the unobservable skills of potential employees and cannot perfectly observe true employee productivity. Employers therefore use knowledge from interactions with others with the same characteristic to make decisions about an individual. A person is therefore not seen as an individual but as a member of the group they belong to. In the absence of reliable individual information, the person will be assumed to have the average productive characteristics of their group. For example, when filling a full-time position, an employer who thinks that women are more likely to take time off for childbearing, will use this information to bias them towards hiring a male worker. Statistical discrimination can lead to self-fulfilling prophecy in that members of the disadvantaged group (women) may invest less in productivity characteristics because they know that is how they are perceived in the labour market in any case (Lang & Lehmann 2011). Continuing with this argument, employers who anticipate career interruptions for women due to family responsibilities will invest less in female on-the-job training.

Studies also document pre-labour market discrimination as a source of the gender wage gap in the labour market, a situation where historical restriction on the type of education or training that women receive, disadvantages them in the type of occupations or careers they participate in. For example, if caregivers or guardians know there is a big chance that their daughter will face discrimination in a certain career, they will not encourage her to pursue that career (Altonji & Blank 1999).

Occupational segregation has also been used to explain the gender wage gap. If women are excluded from some types of jobs considered “male” jobs, then this can result in overcrowding in jobs considered “female” jobs. The overcrowding in turn drives down wages (Bergmann 1974).

Another factor that can affect the gender pay gap is the overall wage structure[20] within the labour market (Blau & Kahn 2000). If there is a rise in the rewards for certain types of individuals or labour market characteristics due to shifts in the labour market demand for these characteristics, whoever is more advantaged in those characteristics will have a wage advantage. Blau & Kahn (1997) discuss the connection between the recent technological change, the decline of unionised manufacturing jobs and the narrowing of the gender wage gap at the bottom of the wage distribution. Historically, women have been underrepresented in manufacturing jobs and are less likely to be unionised, therefore, with the introduction of technology and the increased use of computers, there was a decline in the rewards for physical strength. This shift benefited women relative to men at the bottom of the wage distribution leading to a narrowing of the gender wage gap. The shift however, benefited men relative to women at the top of the wage distribution leading to an expanding gender wage gap at the top of the wage distribution (Blau & Kahn, 1997).

[20] The wage structure is the “array of prices determined for labour market skills and the rewards to employment in particular sectors” (Blau & Kahn 2000, p.6).

3.2.2 Review of Previous Empirical Research on the Gender Wage Gap

Until recently most studies on inequality in South Africa focused on race. However, as labour force participation of women increased, several studies have analysed the gender wage gap either in descriptive studies, mean decompositions or over the entire wage distribution. Below we review some of the studies on the gender wage gap in South Africa. We divide the studies into those that give the magnitude at one point in time and those that compare more than one period and therefore say something about the evolution of the wage gap over time.

3.2.2.1 The Size of the Wage Gap

There is a lot of variation in the type of sample and methodology used in the analysis of the gender wage gap in the South African labour market (see table 15 in the appendix). However, the recurring theme is that the gender wage gap in South Africa persists today (Muller 2009; Ntuli 2007b; Shepherd 2008). There is also an indication of a declining gender wage gap from approximately 29% (Rospabé 2001) or 27% for full time employees (Muller 2009) in 1999, to approximately 18% for full time employees in 2006 (Muller 2009).

Winter (1999) used the 1994 OHS to investigate female labour force participation and the gender wage gap. The study focused on individuals between the age of 15 and 65 years in the formal sector. The author used the log of weekly wages as the dependent variable and states that she dealt with outliers by excluding from the sample any individual whose income fell in the top or bottom standard deviation. This is however peculiar as the study would have lost a lot of observations. Results from the Oaxaca decomposition showed that, on average women earned 87 percent of men's wages in 1994. Additionally, the study found that the gender wage gap was highest within the white population group, where white women earned only 67% of men's wages. The study however reports an almost insignificant wage gap between black men and women. The author finds this result peculiar and interesting and attributes part of it to the female advantage in education for African women.

Although Winter's study limited explanatory variables to education and experience, her results are similar to Hinks (2002) who focused on the OHS 1995 and found that the gender wage gap was highest between white men and women with women earning only 54% of the men's wages. Interestingly, Hinks found that African women enjoyed a 10% wage advantage over men in 1995, a result he also attributes to higher productivity of female workers given their higher average education. Hinks however cautions against the interpretation of this result as the low-pay domestic workers who were predominantly black, and female were underrepresented in the 1995 sample (Hinks 2002).

Similarly, Grün (2004) assessed direct and indirect gender discrimination in the South African labour market using data from the OHS 1995, the OHS 1997 and the OHS 1999. Oaxaca decomposition results for 1995 showed a raw wage gap in favour of African women (log 0.1799) a result the author indicated was unrealistic and postulated that the wage advantage in favour of African women was probably due to incorrect wage figures for men in this year. In chapter two of this thesis we showed that these peculiar results for 1995 can actually be attributable to classification inconsistencies and under sampling of domestic workers in 1994 and 1995.

Rospabé (2001) uses the OHS 1999 to analyse gender inequality and discrimination in labour market outcomes (employment, occupational entry and wages) in the South African labour market. To address the issue of earnings given in brackets, the study utilizes interval regressions instead of a normal OLS regression. Using the Oaxaca decomposition, the author estimates the gender wage gap in 1999 at 29% (0.257 log points) more than half of which she states might be as a result of discrimination. The author also finds a high disparity in occupational distribution and reports that even though South African women have access to high skill occupations, they still find themselves squeezed in a few low-level industries which could also be linked to discrimination in the labour market. Additionally, a third of the gap in access to the labour market (formal and self-employment) could not be attributed to gender differences in human capital characteristics. Disaggregating these results by race, like Hinks, Grün and Winter, Rospabé finds that the gender wage gap is highest between white men and women and estimates it to be at 42%. However, unlike the previous studies, the wage gap between African men and women is high, with men receiving on average 40% higher wages than women, thus giving the impression of an expanding gender wage gap between 1994 and 1999 among the African sub-population. One of our contribution to this literature is to show that after correcting the inconsistent classification of domestic workers in the OHS 1994 and 1995 and using the PSLSD 1993 as a new baseline instead of the OHS 1995, the gender wage gap actually declined over this period.

Most recently Bhorat & Goga (2013), re-examined the gender wage gap in post-apartheid South Africa using data from the September round of the LFS 2007. They apply the re-centred influence function (RIF) decomposition and find that the wage gap is wider at the bottom of the distribution and reduces as we move up the distribution. From the decomposition results they report a gap of 0.632 log points at the 10th quantile and only a 0.072 log points gap at the 90th quantile for Africans. To deal with the bracket earnings problems, the authors apply the 'midpoint' method which has been applied by other studies in the wage literature (Posel & Casale 2006). However, the drawback of this study is that it only focuses on LFS 2007.

3.2.2.2 The Evolution of the Gender Wage Gap over Time

To analyse trends in earnings in the South African labour market, Casale (2004) uses earnings data from the OHS 1995 and the September round of the LFS 2001. Using descriptive analysis, the author examined individual earnings by education level, employment type and occupation and concluded that the gender gap in earnings in the South African Labour market persisted between 1995 and 2001. The author reports that women persistently earn less than men for the same level of education, same type of occupation and employment.

Few studies have investigated the gender wage gap by analysing the entire wage distribution. An exception is Ntuli (2007b) who applied quantile regressions, on the 1995 and 1999 waves of the OHS and the 2004 wave of the LFS to analyse the gender wage gap among Africans over the entire conditional wage distribution. Following Machado & Mata's (2005) bootstrap method, results from this study revealed that in the period between 1995 and 2004, the counterfactual wage gap did not decline over the entire wage distribution and that the wage gap was most severe at the lower quantiles of the wage distribution. Additionally, the paper finds that discrimination had worsened for women in the upper quantiles over this period. Her results also reveal that age (experience) is a significant factor in explaining the gender wage gap but this significance declines as we move higher up the wage distribution. However, results from this study are overshadowed by the fact that the author does not indicate how she treated earnings that were reported in brackets and, she uses 1995 as her base year which recent studies have found to be riddled with peculiarities.

Muller (2009) applying both the Oaxaca-Blinder decomposition and the Juhn-Murphy-Pierce decomposition (Juhn et al. 1993), investigates female part-time employment in South Africa using the OHS 1995 and 1999 and the LFS 2001 and 2006 and focusing only on wage employees. The author reports a declining gender gap within this period contrary to Ntuli (2007b). This contradiction can be attributed to the fact that Muller (2009) dropped the OHS 1995 from the analysis after finding a negative wage gap in this year. Muller attributed the negative wage gap to the under sampling of domestic workers in this year (Muller 2009, p.2) and notes that including the OHS 1995 would have given the impression of a worsening wage gap over time.

Unlike the studies above, Shepherd (2008) conducts an analysis of the wage gap by utilizing 11 years of household surveys (the OHSs and the LFSs) covering the period between 1996 to 2006. Examining only the formal sector and excluding domestic workers and subsistence agricultural workers, the study finds a positive unexplained gap and a negative explained gap among Africans for the period analysed. The analysis also finds a declining overall gap for Africans over the period which the author suspects could be due to the implementation of labour market legislation against unfair discrimination. While this is the first study to recognise the benefit of using multiple consecutive datasets, the distributional analysis focused on Africans only and the study period starts in 1996 two years after the end of apartheid. In addition, the study does not clearly indicate how issues of missing information and outliers were dealt with.

Using this literature, we speculate that the gender wage gap in the South African labour market could have been declining over time.

3.2.2.3 Issues in Estimating the Gender Wage Gap

3.2.2.3.1 Selection

Many studies have tried to control for sample selection using the Heckman two stage model (Hinks 2002; Ntuli 2007b; Shepherd 2008) however the selection coefficient (Lambda) from most of the studies turns out to be insignificant. Ntuli (2007b), analysing the gender wage gap using the OHS 1995 and 1999 and the LFS 2004, included a selection coefficient for the probability that an individual would participate in the labour market and another for the probability that an individual is in employment if they do participate in the labour market. The coefficients were insignificant in most regressions. She concluded that the selection bias was possibly not binding in the formal sector. Shepherd (2008) also chose to ignore selection bias because over the 11-year period (the OHS 1996-1999 and the LFS 2000-2006) analysed, less than half of the coefficients on the female lambdas were significant. Other studies cite difficulty in finding valid instruments that meet the exclusion restriction that is, variables which determine the probability of participation but are not related to wages (Muller 2009). All these studies acknowledge that their estimates may be biased by the inability to correct for selection.

3.2.2.3.2 Specification

There is a lot of similarity in the control variables used in the wage regression in the South African literature. However, there is a debate on whether to use potential experience (calculated as *age - years of schooling - 6*) when actual experience is not available or to use age and its quadratic to proxy for experience instead. Mincer (1974) recommends the use of potential experience as it reflects the fact that people who stay in school longer forgo joining the labour market earlier. Studies of the South African labour market are divided between those that use potential experience (Bhorat & Goga 2013; Winter 1999; Rospabé 2001; Shepherd 2008) and those that use age and age squared (Ntuli 2007b; Hinks 2002). Those that use age and its quadratic follow the argument that due to grade repetition and long spells of unemployment, South Africans do not work continuously after completing their education therefore using potential experience would overestimate the amount of experience obtained (Keswell 2004; Hinks 2002; Keswell & Poswell 2004). However, one can also argue that using the age variable will still produce biased results as this variable does not account for the fact that people who stay longer in school will have fewer years of experience.

Internationally, there is a question of whether one should include occupation and industry dummies in a gender wage gap regression as entry or no entry into these occupations or industries may be because of pre-labour market discrimination and therefore resulting estimates would underestimate the level of discrimination. However, all studies on the wage gap in South Africa include occupational and industry dummies to control for unobservable human capital characteristics that lead individuals to self-select into certain occupations. The exception is Winter (1999) who only had education in years, experience and its quadratic and log of hours worked. The estimates from this regression are likely to be an upper bound as this specification leaves out many important individual characteristics that explain wages such as location, union status and race.

3.2.3 Summary

There is consensus from previous studies that the gender wage gap persists in South Africa. However, there is no consensus about the magnitude of the gap or its evolution over time. This is because studies that have investigated the gap have focused on different points in time, using only a limited set of available data. Moreover, the different data quality issues raised in chapter 2 have not been addressed by these studies making comparability of the results difficult.

In the years after these studies were published, a lot of data work has gone into improving the comparability of household surveys in the Post-Apartheid Labour Market Series (PALMS) (Kerr et al. 2016). In addition, in chapter two we add to these improvements by addressing the inconsistency in classification of domestic work in 1994 and 1995.

There is a need to re-examine the gender wage gap in South Africa using these more robust data over a longer period starting with a better baseline than the OHS 1995 or OHS 1994. By using better data (PALMS 1993-2015) this research contributes to this literature by analysing the gender wage gap over a longer period which helps us isolate trends of the wage gap due to data effects and those due to economic and social changes.

3.3 THE DATA AND METHODS

3.3.1 The Data and Measures

For this chapter we continue to utilize the PALMS dataset discussed in section 1.3. The data used for this analysis has been adjusted to fix the inconsistency in the classification of domestic workers in the OHS 1994 and 1995. We restrict our sample to individuals aged between 15 and 65 years and in wage employment. The restriction to wage earners is to ensure comparability of earnings as there is controversy over self-employment self-reported earnings. By using the earnings variable provided in the PALMS we deal with the issue of missing information and earnings given in brackets, since PALMS contains multiple imputations for earnings using the hot deck method (see Wittenberg 2016b for more details). For this analysis, the explanatory variables include a four-category potential experience[21] variable calculated as *age - years of schooling - 6*, a four category education[22] variable, dummy variables

[21] The four categories of potential experience are constructed as follows: individuals with less than 10 years of experience (0-9 years), those with between 10 years and 19 years of experience (10-19 years), those with between 20 and 29 years of experience (20-29 years) and finally those with more than 30 years of experience (30-59 years).

[22] We construct 4 categories of education as follows: the first category (Primary) includes everyone with between zero and 8 years of education, the second category (Incomplete secondary) includes individuals with more than 8 years of education but less than 12 years of education, the third category (Matric) includes all individuals with 12 years of education and the fourth category (tertiary) includes all individuals with more than 12 years of education.

for being married, unionised, and for whether someone is in the public sector, a 4 category race variable (1=African, 2=Coloured, 3=Indian, 4=White) and a 9 category province[23] variable.

We also include occupation and industry variables. The dataset contains 10 category occupation and industry variables, however, the domestic services sector contains very few men, so to avoid common support issues we combine domestic work with the elementary occupations. For the same reason, we combine the domestic services sector with the services sector.

Like other studies looking at the gender wage gap (Bhorat & Goga 2013; Muller 2009; Grün 2004), we use the log of hourly wage as our dependent variable to allow for the fact that women might spend less time in labour market production due to family responsibilities (Weichselbaumer & Winter-Ebmer 2005). As we do not have hourly wages in our data set, the hourly wage variable was constructed by dividing real monthly earnings by monthly hours. Monthly hours are calculated by multiplying hours worked in the last week by average weeks in a month. For this analysis we took average weeks in a month to be 4.333.

The relationship between schooling and wages in the South African labour market has been found to be convex (Keswell & Poswell 2004). That is, the effect is smaller at lower levels of education and increases with higher levels of education. We account for this non-linearity by including four categories of education levels in our analysis instead of just using years of education.

Marital status is included as a control for productivity. However, while marriage might signal potential increase in productivity for men, it may signal a potential reduction in productivity for women (Blau & Kahn 2016, 2006; Weichselbaumer & Winter-Ebmer 2005). This is because to the employer, being married for men signals 'stability, discipline and motivation' (Rospabé 2001) while for females it signals added non-work responsibilities and less productivity (Weichselbaumer & Winter-Ebmer 2005, p.495). Race is an important covariate in the South African labour market given the apartheid history.

The union status dummy accounts for the fact that union jobs tend to pay higher wages (Butcher & Rouse 2001; Schultz & Mwabu 1998) and they are more likely to be male dominated and thus omitting this variable would overestimate the wage gap (Weichselbaumer & Winter-Ebmer 2005). We also include a

[23] The nine provinces are Western Cape (wc), Eastern Cape (ec), Northern Cape (nc), Free state (fs), Kwazulu-Natal (KZN), North West (nw), Gauteng (gt), Mpumalanga (mpl) and Limpopo.

dummy for whether someone is employed in the public sector because the public sector is an important employer of women in South Africa and studies have found a public sector premium in wage regressions (Heintz & Posel 2008; Bhorat & Goga 2013). Important to note is that in trying to use all waves of data available, a challenge we face is that not all variables are available in all waves. In our case, the union status variable and the public sector variables are not available in some waves. Therefore, to control for the effect of these variables, we run separate regressions for waves where these variables are available.

We considered whether to include occupation and industry dummies in our regression model. If we assume that the type of occupation an individual self-selects into is purely dependent on choice and not due to pre-market discrimination, then inclusion of occupation and industry dummies contributes important information to the model. However, if selection into occupations is due to some form of pre-labour market discrimination, then occupation dummies reduce part of the causal relationship between wages and gender as they are themselves part of the effect (discrimination) we are trying to estimate, and our estimate will be biased downwards. In this case, occupations and industry is a channel through which the gender variable influences wages.

There is a second problem with the inclusion of occupation and industry dummies in the wage regression, that these variables are 'bad controls'. As defined by Angrist & Pischke (2008), bad controls are variables that are themselves outcome variables i.e. they could themselves be the dependent variables of interest. The 'bad control problem' is a different type of selection bias (Angrist & Pischke 2008). To illustrate this, they look at the effect of including both education and occupation dummies in the well-known Mincerian wage regression when estimating returns to education. Education is one of the main determinants of wage, however, it is also one of the main determinants of the type of occupation one finds oneself in (white collar or blue collar). That is, highly educated individuals are more likely to be in white collar occupations and low educated individuals are more likely to be in blue collar occupations. On the chance that an individual with low education is observed in a white-collar occupation despite their low education, it must be the case that this individual is fundamentally different from their other low education peers, probably with very high motivation and or very high innate ability. Therefore, in this context, inclusion of industry and occupation dummies in the wage regression introduces a selection bias component.

Our study considers these two issues, however, as we are unable to model selection into different occupations we run two sets of wage regressions: the first set excludes the occupation and the industry dummies, and the second set includes them. We suggest that our two estimates provide lower and upper bounds (Arulampalam et al. 2007).

3.3.2 Estimating the Gender Wage Gap at the Mean

3.3.2.1 Single Equation Estimation with a Gender Dummy (OLS)

The easiest way of estimating the gender wage gap is to include a female indicator variable in an OLS wage regression with a pooled sample of male and female workers. Regressing log of wages only on the female indicator variable estimates the unadjusted 'raw' gender wage gap[24]. The estimation takes the form of equation (5) below where lnW_i is the log of hourly wages for individual i, α is an intercept, D_i is the female indicator, γ is the unadjusted or the 'raw' gap coefficient, and η_i is the error term.

$$lnW_i = \alpha + D_i'\gamma + \eta_i \tag{5}$$

One can then estimate the adjusted[25] wage gap as per equation (6) where X_i represents the different individual and job characteristics that determine someone's wage and the βs are coefficients on these characteristics and lnW, γ, D are defined as above. ε is an error term and γ in this case captures the adjusted or unexplained wage gap. This coefficient represents 'discrimination' if all the variables important for productivity are controlled for in the regression. However, controlling for all variables related to wages is quite difficult since information on some variables is simply not available in most datasets or not observed.

$$lnW_i = X_i'\beta + D_i'\gamma + \varepsilon_i \tag{6}$$

An important critique of this method of estimating the gender wage gap is that it does not consider gender differences in returns to important characteristics (O'Neill & O'Neill 2006; Goraus et al. 2015). With equation (6) there is an underlying assumption that the returns to relevant characteristics as depicted by the coefficients can be approximated by the average returns for the men and women included in the sample. The problem, however, is that these returns may differ significantly between

[24] The unadjusted wage gap is the effect of being female on wages before controlling for productivity and demographic characteristics.

[25] The adjusted gap, also referred to as the 'unexplained gap', is the wage gap that remains after controlling for observable productivity characteristics

men and women. For example, the returns to marriage as discussed earlier may vary between men and women. The suggested solution around this problem is to estimate separate regressions for men and women and then decompose the absolute wage differential into a component due to differences in productivity characteristics and a component due to differences in rewards to those characteristics (Goraus et al. 2015). This can be achieved by estimating two separate wage regressions and performing the Oaxaca decomposition to determine what proportion of the raw gap is attributed to productivity characteristics and what proportion can be attributed to the differences in rewards. We discuss the Oaxaca decomposition below.

3.3.2.2 Oaxaca Blinder Decomposition (OB)

The Oaxaca Blinder decomposition method (Oaxaca 1973; Blinder 1973) is a counterfactual technique that decomposes the mean wage gap into an explained and an unexplained component. The idea is that if the labour market was fair, two groups with the same labour market productivity should earn the same wage. The technique helps us answer the question, 'How much would women be paid in mean wages if they had the same productivity characteristics as men?' The 'explained gap' is the difference in wages due to observable productivity characteristics between men and women, for example education and experience. The 'unexplained gap' usually referred to as 'discrimination', is the residual due to differences in the economic return on these characteristics depending on whether an individual is male or female.

The Oaxaca decomposition requires that the Mincerian wage equation augmented with other characteristics that have been found in the literature to influence wages be estimated separately for men and women. The simplest way to do this is to run two separate earnings regressions while controlling for other wage determining characteristics such as sector of economy or geographical factors such as urban and rural locations.

Let the log of wages for females and males be determined by equation (7) and equation (8) respectively

$$lnW_f = X_f'\beta_f + \varepsilon_f \tag{7}$$

$$lnW_m = X_m'\beta_m + \varepsilon_m \tag{8}$$

where lnW_f and lnW_m are log wages earned by women and men and X_f and X_m are vectors of human capital and demographic characteristics important for wage determination among men and women respectively and ε_f and ε_m are the respective errors. According to Oaxaca (1973) and Blinder (1973), the difference in mean wages can be defined in equation (9) as

$$\overline{lnW_m} - \overline{lnW_f} = (\bar{X}_m - \bar{X}_f)'\hat{\beta}_m + \bar{X}_f'(\hat{\beta}_m - \hat{\beta}_f) \tag{9}$$

or in equation (10) as

$$\overline{lnW_m} - \overline{lnW_f} = (\bar{X}_m - \bar{X}_f)'\hat{\beta}_f + \bar{X}_m'(\hat{\beta}_m - \hat{\beta}_f) \tag{10}$$

The first terms on the right-hand side of the two equations represent the composition effect or as commonly known, the 'explained wage gap' attributable to differences in quantities (covariates), while the second terms give the wage structure effect or the 'unexplained wage gap' which represents differences in returns on the characteristics (coefficients). A more generalized form of the Oaxaca decomposition equation can be expressed as equation (11)

$$\overline{lnW_m} - \overline{lnW_f} = (\bar{X}_m - \bar{X}_f)'\beta^* + \bar{X}_m'(\hat{\beta}_m - \beta^*) + \bar{X}_f'(\beta^* - \hat{\beta}_f) \tag{11}$$

Where the first term on the right hand side represents the composition effect and the last two terms represent male advantage and female disadvantage respectively.

The debate usually is what β^* should prevail under no discrimination (Oaxaca 1973; Oaxaca & Ransom 1994; Neumark 1988), if $\beta^* = \hat{\beta}_m$ then we end up with equation (9) if $\beta^* = \hat{\beta}_f$ we will end up with equation (10). Alternatively, β^* can be an average of $\hat{\beta}_m$ and $\hat{\beta}_f$ using the sample proportions of men and women as weights. According to Neumark (1988) and later supported by Oaxaca & Ransom (1994), coefficients from a pooled regression over both males and females should be used where $\beta^* = \Omega\beta^M + (I - \Omega)\beta^F$ and Ω is a weighting matrix denoted as $\Omega = (X'X)^{-1}X_m'X_m$ (Oaxaca & Ransom 1994). X in this case is the observation matrix from the pooled male and female sample and X_m is the observation matrix from the male sample.

Important to note is that the magnitude of the composition and the wage structure effects will be dependent on the choice of reference wage structure that is whether $\beta^* = \hat{\beta}_f$, $\beta^* = \hat{\beta}_m$ or $\beta^* = \Omega\beta^M + (I - \Omega)\beta^F$. The wage structure effect (unexplained gap) is highest when the female wage structure ($\beta^* = \hat{\beta}_f$) is used as the reference wage structure and lowest when the male wage structure ($\beta^* = \hat{\beta}_m$) is used. As we are interested in the counterfactual wage for men if they were paid like women, we pick $\beta^* = \hat{\beta}_f$ for this analysis ending up with equation 10. Additionally, this will allow us to compare the OB decomposition results to results from the non-parametric approach by DiNardo et al. (1996).

A limitation of the Oaxaca decomposition is that it is a parametric approach and thus it assumes a particular functional form (linear function) in the earnings distribution which increases the chances of a specification error if the relationship between earnings and the explanatory variables for either men or women is incorrectly specified. This means that the inference regarding the portion that is due to differences in characteristics is likely to be biased (Barsky et al. 2002). A suggested solution to this problem is to carry out a reweighted decomposition at the mean similar to the one formulated by DiNardo et al. (1996). The choice of $\beta^* = \hat{\beta}_f$ enables us to compare results from the OB decomposition to results from the Dinardo, Fortin and Lemieux decomposition where we also estimate the counterfactual wage men would be paid if they were paid as women.

An advantage of the OB decomposition over the OLS regression with a female indicator is that the detailed decomposition results from the OB give an estimate of the contribution of each explanatory variable to the explained and unexplained gap. However, interpreting the contribution of each explanatory variable to the unexplained gap is not straightforward. This is because when a model includes categorical variables with more than two categories, the detailed decomposition results are not invariant to the choice of the base category. This is the well documented identification problem referred to as the 'omitted category' problem (Fortin et al. 2011; Oaxaca & Ransom 1999; Jones & Kelley 1984; Yun 2005).

Yun (2005) and Gardeazabal & Ugidos (2004) have proposed a solution to this identification problem by suggesting estimation of some form of 'normalized' regression equations where the restriction that all of the dummy variables' coefficients sum to zero is imposed. However, there is no consensus that this is the ultimate solution. For example, Fortin et al. (2011, p.41) point out that "the normalization proposed may actually leave the estimation and decomposition without a simple meaningful interpretation". Additionally, due to the lack of a standardized rule on the categories to be omitted, the results from any such normalization will probably be sample specific therefore negating chances for comparability of estimates across studies (Fortin et al. 2011; Oaxaca 2007). As a solution to this issue, Fortin et al. (2011)

opine that, researchers should avoid 'automatic normalization' but instead choose 'reasonable' and 'interpretable' base categories.

As there is no agreed upon reference group from the literature, this study uses the single elementary worker, with less than 10 years potential experience, from the Western Cape and with primary education or less in the manufacturing sector as its omitted group. This is because many employed women are concentrated in elementary and domestic work which require lower skills than other occupations.

3.3.3 Estimating the Gender Wage Gap across the entire Wage Distribution

3.3.3.1 The Dinardo Fortin and Lemieux (DFL) Re-weighting Approach

The DFL methodology is a generalization of the Oaxaca Blinder decomposition where the coefficients or returns to characteristics in the Oaxaca Blinder decomposition are now thought of as the conditional wage distribution. However, whereas in the Oaxaca decomposition we construct a mean counterfactual, in the DFL we are analysing the distributional counterfactual. Moreover, the DFL is semi-parametric, therefore no particular functional form of the wage distribution is assumed. From the program and evaluation literature, Hirano et al. (2003) show that the re-weighting estimator is asymptotically efficient. More recently, in a review of decomposition methods in the labour market, Fortin et al. (2011, p65) recommend this method for aggregate decompositions as it provides consistent estimates of the wage structure and composition effects for any distributional statistic of interest.

In this section we are following the exposition of the method in DiNardo et al. (1996, p.1010). The basic idea of the DFL is to construct a counterfactual distribution of wages of one group, in our case women, by replacing the productivity characteristics with those of another group, in our case men, using a re-weighting factor. The aim is to answer the question 'What would the distribution of wages for men be if they were paid as women?' We then compute the aggregate composition and wage structure effects over the conditional wage distribution using the counterfactual wage distribution.

DFL view each individual observation in a given wage distribution as a vector w_i (w_i,x,j) composed of the wage (w_i), a vector of individual attributes (x) and a group subscript (j), where ($j = F, M$). We express the observed distribution of wages as

$$f(w) = \int g(w|x)h(x)dx \quad (12)$$

Where $g(w|x)$ is the conditional distribution of wages given the characteristics x and $h(x)$ is the marginal distribution of x (observed productive characteristics). The observed distribution for the female group is thus given by

$$f(w|j = F) = \int gF(w|x)h(x|j = F)dx \quad (13)$$

Where $gF(w|x) = g(w|x, j = F)$ is the female conditional distribution of wages given the characteristics x and $h(x)$ is the distribution of x (productive characteristics) for females. The counterfactual distribution where we ask what the wage distribution of males would be if they were paid like women, is given by

$$g_F^c (w|j = M) = \int gF(w|x)h(x|j = M)dx \quad (14)$$

where $g_F^c (w|j = M)$ is the counterfactual distribution of wages observed for men if they were paid according to the female wage distribution. This is under the strong assumption that the female distribution of wages conditional on characteristics does not depend on the distribution of characteristics for females. The hypothetical density can be written as

$$g_F^c (w|j = M) = \int \Psi_x(x)gF(w|x)h(x|j = F)dx \quad (15)$$

where $\Psi_x(x)$ is the re-weighting function and is defined as

$$\Psi_x(x) = \frac{Pr(j = M|x)}{Pr(j = M)} \frac{Pr(j = F)}{Pr(j = F|x)} \quad (16)$$

$\Psi_x(x)$ is a function that maps the male distribution of characteristics onto the female distribution. It re-weights the female density so that observations that are more likely for males than for females are weighted up, and observations that are less likely are weighted down. It can be estimated from the data by using a standard probability model on data where males and females are pooled together. In this analysis $\Psi_x(x)$ is estimated using a logit model for the probability of being female relative to the probability of being male.

The remainder of the expression under the integral sign in equation (15) is just the observed joint distribution of wages and characteristics for females. DFL suggest estimating $g_F^c (w|j = M)$ using standard non-parametric kernel density methods using actual female earnings distribution $gF(w|x)h(x|j = F)$ but re-weighting it with $\Psi_x(x)$. However, since our interest is simply the counterfactual earnings at each quantile, we do not estimate kernel densities, we calculate the gender gap at every quantile using the re-weighted actual distribution of wages for females.

Finally, the overall difference in the decomposition is calculated as

$$\Delta_O^{g(w)} = g_M(w) - g_F(w) = (g_M(w) - g_F^c(w)) - (g_F^c(w) - g_F(w)) \quad (17)$$

The component $g_M(w) - g_F^c(w)$ of equation (17) is the wage structure effect. In this case, since men and women are made to have the same distribution of covariates, the observed difference in wages must be due to the difference in the wage structure (Fortin et al. 2011). The second component of equation (17), $(g_F^c(w) - g_F(w))$, is referred to as the 'explained' component or the composition effect. This is because the assumption is that the difference in wages is solely due to the difference in productivity characteristics between men and women since the wage structure is identical for men and women. This component gives us the aggregate contribution of all the covariates to the overall gender wage gap at different points on the wage distribution.

3.3.3.2 Unconditional Quantile Regressions

A limitation of the DFL approach is that there is no straightforward way to perform a detailed decomposition of the wage structure and composition effects (Fortin et al. 2011, p.65). However, in the case of the composition effect, an extension of the DFL which involves sequentially adding explanatory variables to the probability model used to calculate $\Psi_x(x)$ has been applied in the literature (Altonji et al. 2012; Antecol et al. 2008; Kassenboehmer & Sinning 2014; Baron & Cobb-Clark 2010). The limitation of the sequential method, however, is that the contribution of a particular variable is path dependent, that is, the results depend on the order in which the variable was introduced (Fortin et al. 2011, p.80).

As a solution to the above limitation, Firpo et al. (2009) (henceforth FFL) developed a methodology for estimating the effect of individual characteristics on the unconditional wage distribution using re-centred influence functions (RIF). As our interest is to attribute changes in the wage distribution (w_i) to the effect of individual covariates, FFL show that one is able to apply the Law of Iterated Expectations (LIE) and the result that $E[RIF(W; q_\tau)] = q_\tau$ to retrieve the unconditional distribution of the dependent variable. This result then allows one to perform Oaxaca Blinder type decompositions of the wage gap where in this case, the dependent variable (w_i) is replaced by the *RIF* of that quantile.

According to FFL, for any statistic ν, a functional $\nu(F_W)$ can be defined for the unconditional distribution $F_W(W)$. Using the LIE, they show that the unconditional quantile partial effect (UQPE) is defined as

$$UQPE \neq \beta \ (\nu) = E[\frac{\partial E[RIF(W,\nu)|X]}{\partial x}] \tag{18}$$

where RIF is the re-centred influence function defined as the sum of the original statistic and the influence function i.e. $RIF(w; F_W) = \nu(F_W) + IF(w, F_W)$.

An influence function (IF) is a measure of the influence of an individual observation on a distributional statistic. Further, they show that if q_τ is the τ^{th} quantile of the unconditional distribution of W, the influence function (IF) is defined as

$$IF(w; q_\tau) = \frac{\tau - 1\{w \leq q_\tau\}}{f_W(q_\tau)} \tag{19}$$

where $1\{.\}$ is an indicator function, $f_W(.)$ is the density of the marginal distribution of w, and q_τ is the population τ *−quantile* of the unconditional distribution of w. The re-centred influence function is then calculated by adding back the original statistic to the influence function. Consequently, $RIF\ (w; q_\tau)$ is equal to

$$RIF(w; q_\tau) = q_\tau + \frac{\tau - 1\{w \leq q_\tau\}}{f_W(q_\tau)} \tag{20}$$

Equation (20) shows that the RIF for a quantile is a function of an indicator variable $1\{w \leq q_\tau\}$ for whether the wage is smaller or equal to the quantile q_τ. The first step to estimating the effect of a change in explanatory variables on an unconditional quantile is estimating the conditional $RIF\ (w; q_\tau)$. This is done by computing q_τ and f_W and then regressing the estimated $\widehat{RIF}\ (w; q_\tau)$ on the individual covariates. q_τ is estimated as the sample τ^{th} quantile whereas $f_W(q_\tau)$ can be estimated non-parametrically using kernel density estimation. Then for each observation, we estimate the $RIF\ (w_i; q_\tau)$ by plugging in the estimates $\hat{q}_\tau$ and $\hat{f}_W(\hat{q}_\tau)$ into equation (20). In this study we report the gender wage gap at the 10th, 50th and the 90th percentile.

The advantage with knowing the effect of an individual covariate on the wage distribution at a particular quantile is that different covariates are important at different parts of the wage distribution and different policy interventions will impact different parts of the wage distribution. For example, minimum wage legislation could be more binding on the lower end of the wage distribution. An advantage of the FFL decomposition over sequential DFL and quantile regressions is that the results are not path dependent.

For identification, as is the case in the program evaluation literature the assumptions of ignorability and common support must hold (Fortin et al. 2011, p.16). Ignorability is the assumption that after controlling for observed explanatory factors the distribution of the unobserved variables in the wage determination is the same across men and women. The common support assumption requires that there is no covariate where only members from one group have values. That is, $0 < Pr(j = M|x) < 1$. To ensure that the common support assumption is not violated, especially in the case of domestic work where the probability of males being domestic workers is very low, we have combined domestic workers with elementary occupations.

3.4 MEASUREMENT ISSUES

An important limitation to note is that the decomposition methods we have discussed produce partial equilibrium results (i.e. they assume the invariance of the conditional distribution) (Fortin et al. 2011). That is, they work under the strong assumption that the distribution of wages conditional on characteristics $g(w|x)$ can stay constant if the distribution of the Xs changed. The assumption is that, for instance, the wage distribution of women does not change as the distribution of skills changes. However, in reality this is not usually the case, we might assume that the relative scarcity of various types of education will affect the returns to education. An additional limitation of the RIF methodology is that in cases where the variables are prone to heaping, on certain values such as wages, the performance of RIF regression methods is dependent on the kernel density estimate of f_W which in turn will be determined by the choice of bandwidth (Fortin et al. 2011, p.77). To check for sensitivity to this, we compare our RIF results against results from the DFL aggregate decomposition which is not affected by heaping since the re-weighting factor depends on the group membership and not on the distribution of wages.

3.4.1 Sample Selection Bias

Our sub sample only includes individuals who reported positive earnings and therefore is likely to be non-random. Moreover, in South Africa where the rate of unemployment is high, our sample may not be a representative sample of the working age population. Traditionally, controlling for selection bias involves estimating a probit model for labour force participation and then including the estimates (the inverse mills ratio commonly referred to as lambda) from the probit model in the wage regression as covariates. This is the well-known Heckman two-stage selection model (Heckman 1979).

The procedure however requires presence of valid instruments that are correlated with labour force participation and with employment but are not correlated with earnings. These instruments are referred to as exclusion restrictions and are in practice hard to find, as in the case of this study. According to Puhani (2000), the lack of appropriate exclusion restrictions may generate collinearity issues resulting in unreliable coefficients and inflated standard errors. The procedure is also valid only if the regression errors are normally distributed.

Important to note is that in the South African context, selection into employment is more complicated than a two-stage Heckman selection model would suggest. The very high unemployment rates (which are higher among women than among men) mean that a sizeable share of labour force participants is not employed. That is, there may be selection first among working age men and women into labour force participation and there could be selection among labour force participants into employment. Studies that have controlled for selectivity bias reported that the coefficients for lambda were mostly not significant (Ntuli 2007b; Shepherd 2008; Hinks 2002).

Considering this we do not control for selectivity but acknowledge that our results are likely to be biased and the direction of the bias cannot be known *a priori*. However, descriptive analysis comparing wage employed and unemployed women in terms of age, education and marital status reveals that, in the South African labour market, wage employed women are on average older than unemployed women. This result is consistent with the high youth unemployment problem in South Africa.

Education trends show that wage employed women have on average more years of education. There is a higher proportion of tertiary educated and a lower proportion of high school drop outs (with less than matric) among wage employed women. There is however, no statistical difference between wage employed and unemployed women in the proportion with primary education or less or among those with matric.

In terms of marital status, wage employed women are more likely to be married than unemployed women. These summary statistics show that there are possible selectivity issues, and the direction of the bias is difficult to tell (see figures 36-42 in the appendix).

3.5 RESULTS

3.5.1 Descriptive Statistics

To better understand the results from our analysis we discuss the covariates in our model in this section. A descriptive analysis of the covariates shows that wage employed women have slightly more education than wage employed men and that for both men and women the average level of education of wage employed individuals has been on the rise since 1993. Given the human capital theory that attributes wage differentials to differences in human capital characteristics, women having more education implies a wage advantage for women and therefore a narrowing of the gender gap in wages. The proportion of married men and women has been declining over the period 1993-2015 however the proportion of men in our sample that report being married is higher than the proportion of married women. The decline in marriage rates signals the convergence of career patterns and education patterns of men and women over time. For all the years analysed, men on average recorded more hours of work. In 2015, men recorded on average 45 hours of work per week while women recorded 40 hours of work per week. We find that until 1999, employed men were on average older than women but as of 2000, women's average age has surpassed that of men.

Before 2000, men had on average more potential experience than women however this changed in 2003 where it seems like women have more experience than men. Bhorat & Goga (2013) suggest that this could be because our potential experience variable was calculated using the age variable and employed women on average seem to be older than men especially after the changeover period between the OHS and the LFS. For men the average potential experience has been declining over time. Employed men on average had 22 years of experience in 1997 which dropped to an average of about 21 years in 2015. In contrast, for women, average experience increased from an average of almost 21 years in 1994 to an average of almost 22 years in 2015.

The data shows that union membership has been stable in the South African labour market even though there was a slight increase in union status between 2004 and 2010 and a decline after that for both men and women. 38 percent of employed men reported being in a union in 1997 as opposed to only 30 % of employed women. In 2014, this figure stood at 31 percent for men and 26 percent for women. On the other hand, more women report working in the public sector as compared to men. In 2014, 24 percent of wage employed women reported working in the public sector compared to only 17 percent of wage employed men.

Interestingly, the gender composition in different occupations and industries has not changed much over time. However, there is a slight increase of women managers and legislators which should have a negative effect on the gender wage gap. There is also a slight increase of women in the agricultural and mining sectors (see figure 1).

3.5.2 Single Equation Estimation with a Female Indicator Variable (OLS)

The OLS wage regression showed that all the variables have the expected signs (results not shown here). Wages are positively correlated with being in a union, working in the public sector and being married. Wages also increase with the amount of experience and level of education.

Coefficients for the adjusted and unadjusted gender wage gap are plotted in figure 13. The unadjusted (raw) gap is the effect of being female before controlling for any wage related characteristics. It is simply the coefficient on the female indicator variable from equation (5). The adjusted gap, also referred to as the unexplained gap (Oaxaca 2007), is the effect of being female after controlling for wage related characteristics. In this case it is the coefficient of the female indicator variable from equation (6).

Model 1 depicted by series A shows the estimate of the gender wage gap after controlling for race, education, marital status, potential experience and province. Model 2 depicted by series B shows the adjusted wage gap when we add industry and occupation dummies to the covariates in model 1 whereas model 3 depicted by series C is the adjusted gap when we add a union dummy and a public-sector dummy to model 2. The union and public-sector variables are not available for all the waves and therefore we have only plotted the results for the years where these variables are available.

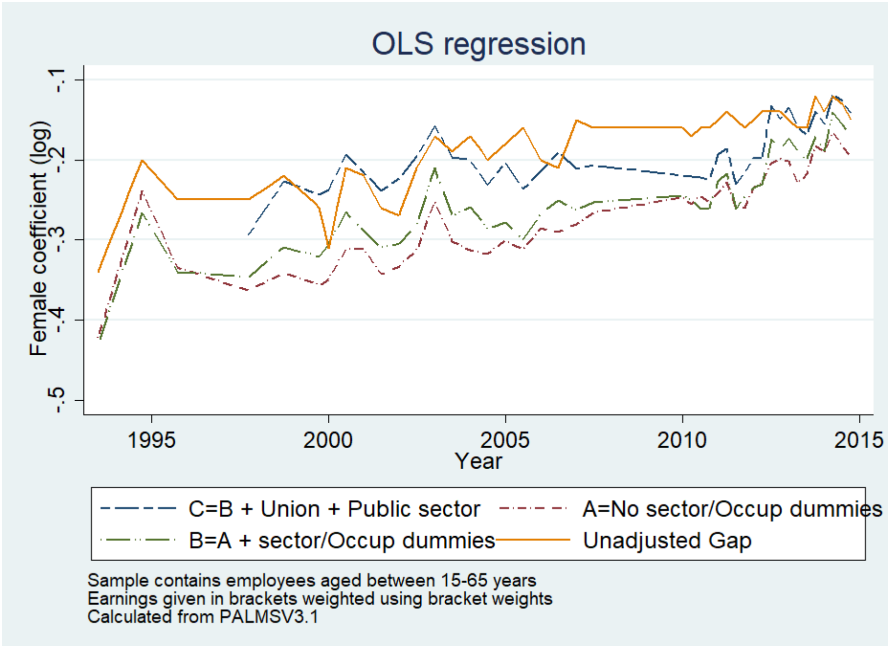


Figure 13: The coefficient on the female indicator variable with and without controls
Notes: Omitted groups: African, single with primary education and with less than 10 years of experience, from the Western Cape, in an elementary occupation in the manufacturing sector, not in a union and in the public sector. A shows the estimate of the gender wage gap after controlling for race, education, marital status, potential experience and province. B shows the adjusted wage gap when we add industry and occupation dummies to the covariates in A whereas C is the adjusted gap when we add a union dummy and a public sector dummy to B.

From the figure we see that the unadjusted gap declined from -0.34 log points in 1993 to -0.15 log points in 2014 whereas the adjusted wage gap (from model 1) declined from -0.420 log points in 1993 to -0.19 log points in 2014. We notice that for model 1 and 2 the adjusted wage gap is always bigger in absolute value than the raw wage gap that is, controlling for gender differences in human capital and demographic characteristics only increases the gender wage gap instead of reducing it. The implication here is that employed women in the South African labour market have an advantage in terms of labour market endowments and thus observable characteristics cannot explain the gender wage gap.

Inclusion of sector and occupation dummies in the wage regression leads to a reduction of the adjusted gender wage gap by about 4.5% from -0.36 to -0.35 log points in 1997 and by 15% from -0.20 to -0.17 log points in 2014. However, the adjusted wage gap still remains more than the unadjusted wage gap meaning that selection into industries cannot explain the gender wage gap. The implication here is that if selection into occupations is fuelled by pre-market discrimination, then the inclusion of occupation and industry dummies underestimates the discrimination estimate.

Series C[26] shows that inclusion of union and public-sector dummies does not qualitatively change the result above. The adjusted gap is less than the unadjusted gap but not significantly so and only for the period between 1998 and 2003. For the rest of the period the adjusted gap is still greater than the unadjusted gap though less in magnitude than the adjusted gap for model 1 and model 2 showing that failure to include these variables gives an upper bound of the unexplained gap.

As discussed in section 3.3.2.1 however, an important critique of the pooled regression with a gender dummy is that it does not take into account the fact that returns to characteristics may differ significantly between men and women. In the following section, we present results from the Oaxaca decomposition where two separate regressions for men and women are estimated and the absolute wage differential decomposed into a component due to differences in productivity characteristics and a component due to differences in rewards to those characteristics.

3.5.3 Oaxaca Decomposition Results

The decomposition is carried out with the user written program Oaxaca by Jann (2008) in Stata. The program runs the earnings regressions for men and women separately, computes the means and the elements of the decomposition along with standard errors that reflect the fact that both the coefficients and the mean values of the covariates are being estimated. We report results for the decomposition where the female wage structure (female coefficients) is used as the reference wage structure. Important to note however, is that using either the male wage structure or coefficients from the pooled model does not alter our results qualitatively.

In figure 14 we present the overall (unadjusted) gap, the explained and the unexplained gap from our analysis and include some results from the South African literature for comparison. The overall wage gap variable gives the average wage differential between men and women whereas the explained component is the mean increase in women's wages if they had the same characteristics as men and the unexplained component is the part of the gap that cannot be explained by differences in characteristics (it is the difference in returns to observable characteristics).

[26] For this series, we excluded OHS 1994, OHS 1995 and OHS 1996 as the Public sector variable was not available in these waves

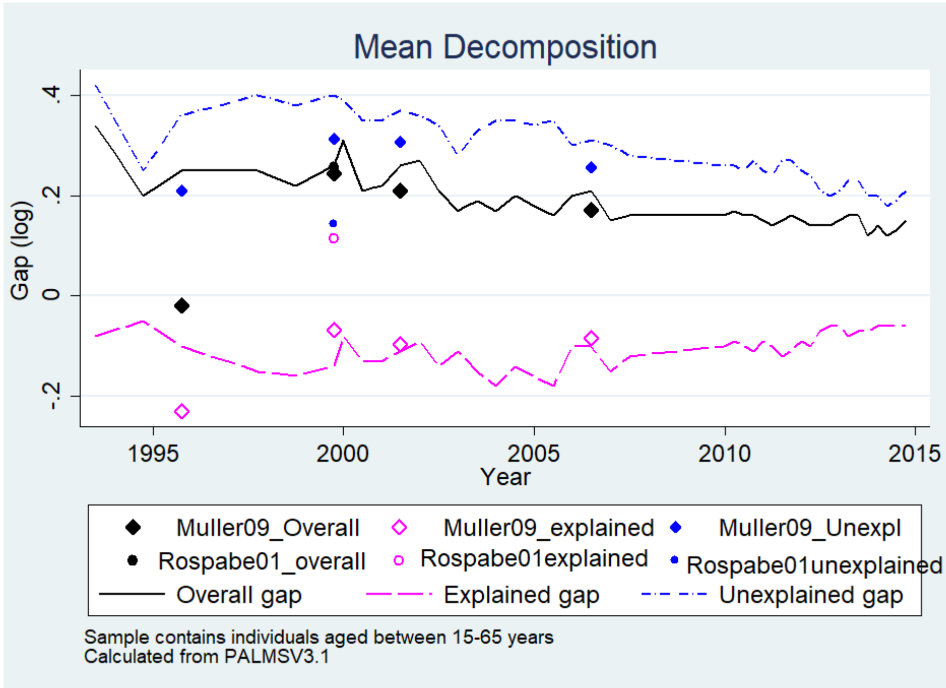


Figure 14: Oaxaca Decomposition Results: Without Sector and Occupation dummies

Notes: Omitted groups: African, single, with primary education and with less than 10 years of experience, from the Western Cape. Muller09 represents results from Muller (2009) whereas Rospabe01 represent results from Rospabé (2001).

Overall, the results show that both the total unadjusted gap and the unexplained gap have been declining since 1993 although the gaps are not going to zero. The trend in figure 14 suggests that the decline of the overall gap at the mean is due to the decline of the unexplained gap (wage structure effects). We would expect that given the implementation of the Employment Equity Act in 1998 which gave way to the enforcement of affirmative action, the unexplained gap would be tending to zero. The fact that the unexplained gap shows a declining trend after 1998 suggests that labour market legislation may have had some effect on the gender wage gap. However, there seems to be some sort of stagnation of the decline of the total unadjusted gap around 2006 where the gap seems to be oscillating around 0.16 log points. The unexplained gap is positive and significant at the 5% level of significance whereas the explained gap is negative and also significant. The negative and significant explained gap suggests that improvement of female human capital characteristics will not help in narrowing the gender wage gap as women already have an advantage in these characteristics.

The gender wage gap can be attributed to differences in human capital characteristics (positive explained gap) or to differences in returns to human capital characteristics (positive unexplained gap). The negative explained gap means that the human capital characteristics cannot explain the gender wage gap and therefore we must look at the wage structure effects (unexplained gap). The persistent gap and the fact that human capital characteristics cannot explain the gap suggests that the gap at the mean may be a manifestation of what is happening in other parts of the wage distribution. It could be the case that different types of labour market legislation affect different parts of the wage distribution differently.

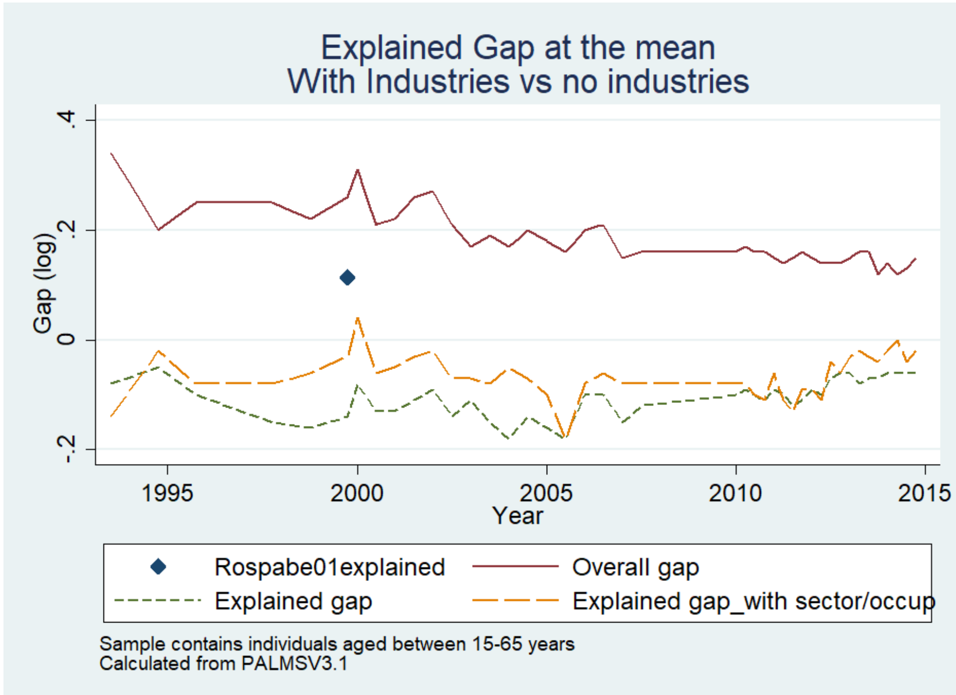
Our results are comparable to some of the results in the literature. For example, except for the OHS 1995, the overall gap is similar with findings from Muller (2009) and Rospabé (2001) as shown in the figure above. The result for 1995 from Muller (2009) however shows that there is value addition in utilizing all available data for any trend analysis. Studies that used the OHS 1995 as a baseline year reported an increase in the gender wage gap between 1995 and 2006 which is clearly not the case from figure 14.

We also note that contrary to Winter (1999) who reports that in 1994 women earned 87% of men's wages, our estimate for the gender wage gap in 1994 was 0.2 log points (approximately 22%) which means that according to our results, women earned only 78% of men's wages this year. These differences are related to the inconsistency in the classification of domestic workers in 1994 and 1995. This stresses the point that data quality issues are important.

The unexplained and explained gaps from our study and Muller's (2009) seem similar but we note that results from Rospabé (2001) are slightly different. This difference could be due to differences in covariates between our studies. In addition to the controls in our model, she also controlled for tenure and whether someone was employed in the formal or in the informal sector. However, as we show in figure 15b the closest estimate to Rospabé is from the model where we include union, public, occupation and sector dummies where we get an explained gap of 0.03 log points against Rospabe's explained gap of 0.114 log points and an unexplained gap of 0.22 log points against the author's 0.144 log points.

Figure 15 presents results for the explained and unexplained gaps from different specifications. The explained gap is negative and mostly significant throughout the series.

(a) Explained gap: Including vs Excluding sector and occupation dummies



(b) Unexplained gap: Including vs Excluding sector and occupation dummies

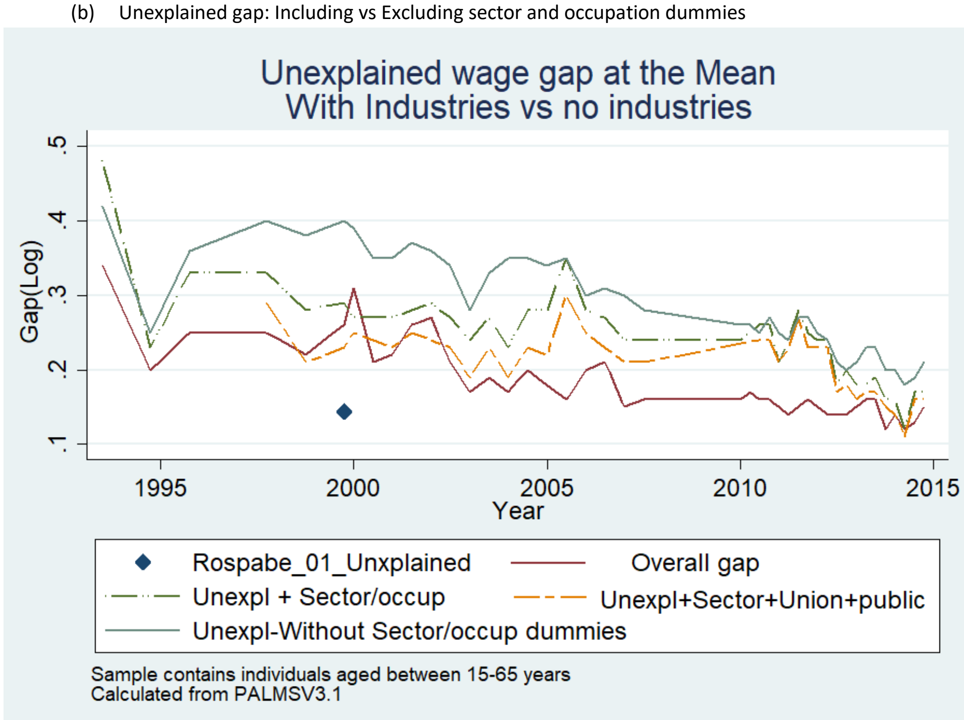


Figure 15: Oaxaca Decomposition Results

As noted in section 3.3.2.2, a limitation of the Oaxaca decomposition is that it is a parametric approach and thus it assumes a linear relationship between earnings and the explanatory variables for both men and women. If, however this relationship is not linear, the explained gap is likely to be biased (Barsky et al. 2002). Below we compare results from the OB decomposition to results from the re-weighting method by Dinardo, Fortin and Lemieux. Our results (see figure 16) show that results from DFL are comparable to results from OB. The main conclusion from this comparison is that the trend of the explained gap and unexplained gap is the same regardless of the methodology used. The two decompositions show that the unexplained gap is positive and persistent whereas the explained gap is negative throughout the series.

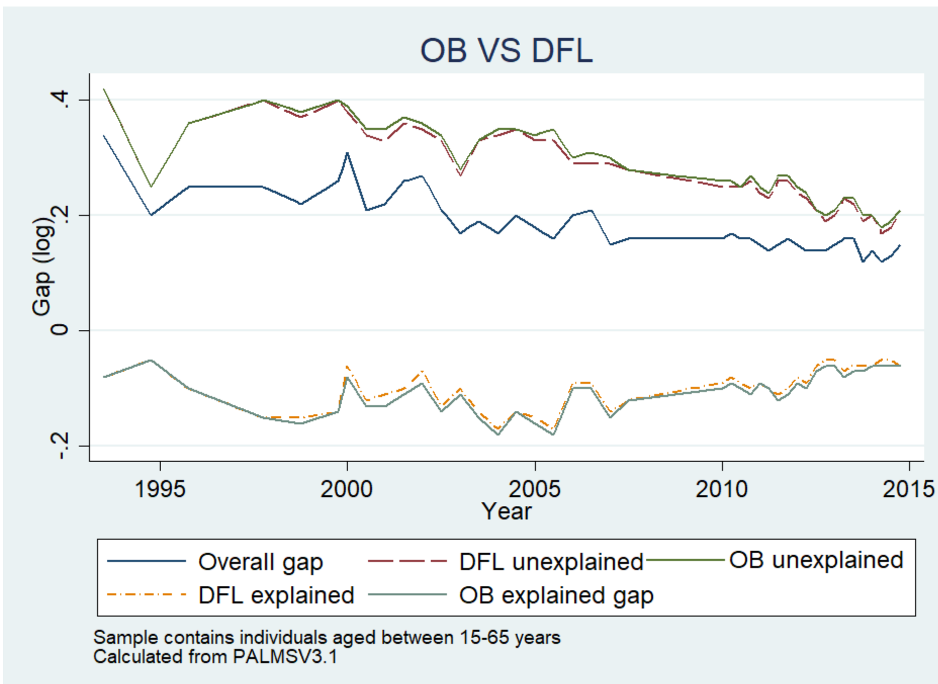


Figure 16: Comparing OB with DFL re-weighting at the mean

3.5.4 Distributional Analysis

Figure 17 shows that changes in the gender gap are heterogeneous across the wage distribution. We see that the female-male wage ratio is highest at the 90th percentile and lowest at the 10th percentile. There has been an increase in the female-male wage ratio over time and over the entire distribution. However, the increase in the wage ratio at the 10th percentile and the median have been much more conservative than at the 90th percentile. The female-male wage ratio increased from almost 0.7 in 1993 to 0.9 in 2014 at the 90th percentile with some fluctuations in between while the wage ratio at the median modestly increased from 0.7 in 1993 to about 0.8 in 2014 with fluctuations in between.

A drastic decline in female wages at the 10th percentile led to a dip in the female-male wage ratio in the OHS. There is especially a sharp drop in the female-male wage ratio at the median and 10th percentile that coincides with the changeover from OHS to LFS. It is documented in the literature that the increase in labour force participation of women during this period was not due to a demand pull but due to women being "pushed" into the labour market because of "economic needs" (Casale & Posel 2002; Casale 2004). As a result, there was an overcrowding of women in low paying occupations which may have pushed wages down even further. On the contrary, the graph shows that there was no contraction of female-male earnings ratio at the 90th percentile in fact this ratio seems to have improved during that period. The trends in the figure stress the need to analyse the changes in the wage gap across the entire wage distribution.

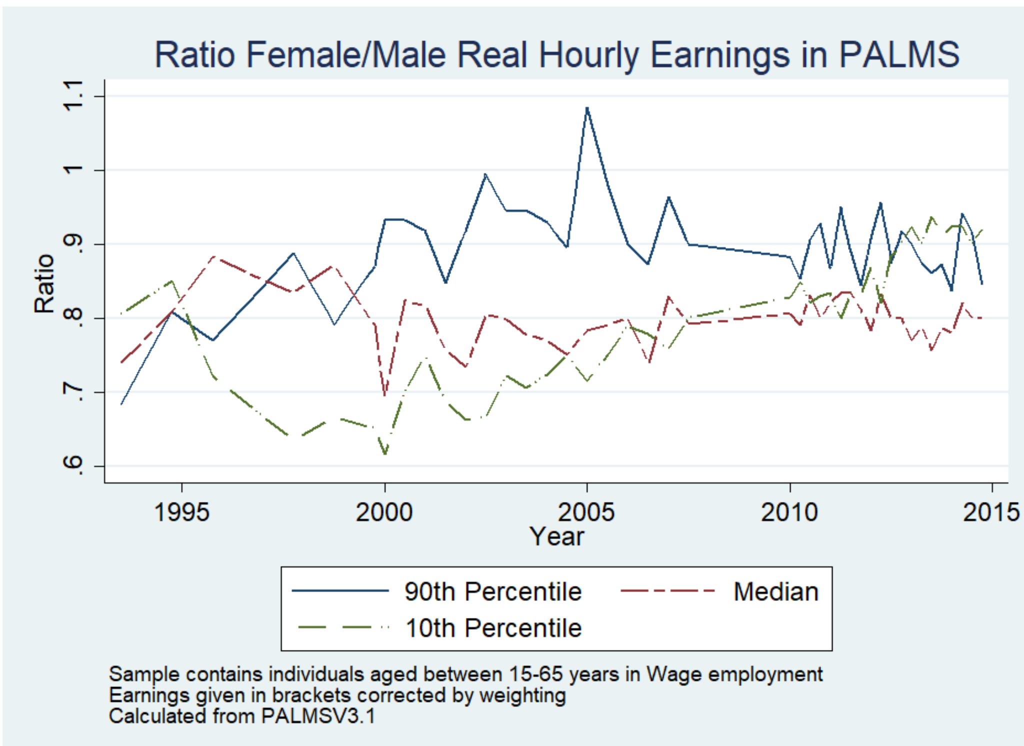


Figure 17: Female-Male Rand Earnings ratio in PALMS

3.5.4.1 D-F-L Aggregate Decomposition

Below we present results for the aggregate decomposition from the DFL methodology. To perform the decomposition, we first estimated a logit model to recover the probability of being male in our sample. However, as Fortin et al. (2011) document, there is some evidence from the program evaluation literature that re-weighting can have undesirable properties when there is a problem with common support. This is because the re-weighting factor will be very large if the probability of an individual being observed in one group is very close to 1 (Fortin et al. 2011, p65). In table 14 we present results from the logit model used to estimate the probability of being male ($Pr\ (M\ =\ 1|X)$) in the pooled sample of men and women. Included in the vector of characteristics (X) is a four-category potential experience variable, four category education variable, a dummy variable for married, a race category variable and a 9-category province variable. We also include sector and occupation dummies.

For better presentation of the table, we have excluded some years[27] and some variables such as the province dummies (these are available at request). The results show that men are more likely to be married than single, more likely to be in professional occupations relative to elementary occupations and less likely to be clerks relative to being in elementary occupations. They are also more likely to be in the mining, construction and transport sectors relative to being in the manufacturing sector. Although in some years some coefficients are quite low, these are not significant.

In figure 18, we consider each percentile of the distribution and show that the effect of characteristics and returns to those characteristics is different at different parts of the wage distribution. For each wave the figure shows that the raw gap rises in the lower part of the wage distribution, peaks between the 20th and the 30th percentile then starts declining. The gap is lowest between the 70th and 90th percentile and then it rises again depicting a 'Sticky floor' effect for employed women in South Africa (Ntuli 2007b; Bhorat & Goga 2013). Over time the peaks at the lower part of the wage distribution seem to flatten as the gender wage gap declines in contrast with the very top of the wage distribution where the gap seems to be expanding especially in the period after 2006 giving an indication of the presence of a 'glass ceiling effect' as well in the South African labour market.

The figure shows clearly that the wage structure effect (unexplained gap), is mostly larger than the overall wage gap especially towards the middle and upper part of the wage distribution. Moreover, as we move up the wage distribution and over time this effect becomes bigger as a percentage of the total gap. This suggests that even though the wage gap is wider at the bottom, women at the top of the distribution face more discrimination. Ntuli (2007b) looking at the African sub sample and using quantile

[27] The reported results are similar to the excluded ones.

regressions arrives at the same conclusion with data from the OHS 1999 and the LFS 2004. This result follows from the idea that at the top of the wage distribution, human capital characteristics such as higher education are more important and as we shall discuss under the detailed decomposition, in the 90th percentile, men receive better rewards for education despite the fact that women have more of it hence the positive and expanding unexplained gap at the top. The explained gap is small and mostly negative and it becomes more negative as we move up the distribution and over time. The explained variables seem to have some importance at the bottom of the wage distribution where we see a small but positive explained gap for example in 1999 and 2004.

The last sample (QLFS 2014:4) in figure 18 however looks anomalous because it shows a very negative wage gap in favour of women at the very bottom which does not seem plausible. The anomaly is most likely due to data quality issues regarding the earnings variable in the most recent QLFS and requires further investigation. In a recent paper investing public sector wages and employment, Kerr & Wittenberg (2016) report that the public sector premium seems to be anomalous after QLFS 2012. They attribute this anomaly to imputations done on the QLFS earnings variable by Statistics South Africa.

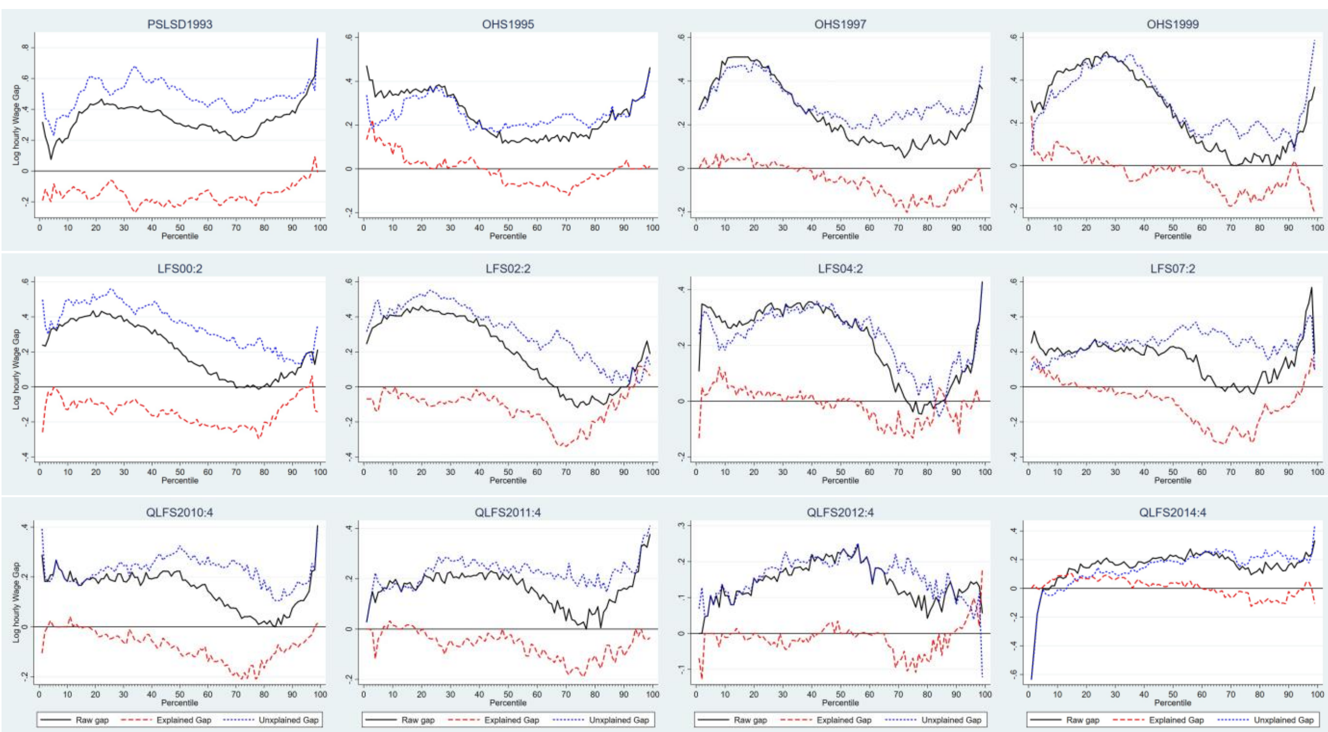


Figure 18: DFL decomposition aggregate gender wage gap across quantiles by wave
Source: Own calculation from PALMSV3.1

3.5.4.2 RIF Decomposition Results: Aggregate Decomposition

Below we present results from the aggregate decomposition using the RIF[28] methodology. Figures 19, 20 and 21 plot the evolution of the gender wage gap at the 10th, 50th and 90th percentile. Figure 19 shows that the overall (total unadjusted) gap at the 10th percentile widened in the beginning of the series from 0.21 log points in 1993 to 0.49 log points in 2000 and has been declining since to about 0.07 log points in 2014. The unexplained gap is positive throughout the series however, it has been declining over time. The explained gap is small as a percentage of the overall gap and is positive after the OHS 1997. The positive explained gap is due to women being concentrated in low paying occupations and industries such as domestic work and elementary occupations. These industries are also less likely to be unionised contributing to the positive explained gap. The explained gap however declined to almost zero after 2006, which is an indication that human capital characteristics between men and women at the bottom of the wage distribution became similar over time.

The widening of the overall gap in the early 1990s was due to a fall in women's wages during this period possibly due to increase in female employment in low paying industries. The trend of the unexplained gap suggests that the decline of the gender gap at the 10th percentile is mostly due to wage structure effects. More so, the timing of the decline of the gap coincides with the implementation of sectoral minimum wage laws for low earning sectors starting with minimum wage legislation for contract cleaners in 1999, followed by the sectoral minimum wages for domestic workers in November 2002 and for Agricultural workers in March 2003. Studies (Hertz 2005; Bhorat et al. 2013) show that there was a substantial increase in wages in the domestic services sector and the agricultural sector as a result of the minimum wage legislation.

The worker at the 10th percentile is more likely to be female, unskilled, in the elementary or domestic work occupation and in the agricultural or retail industry. Therefore, increasing wages in these sectors is bound to improve the position of women. We partly attribute the decline in the gender wage gap exhibited in figure 19 to this increase in wages at the bottom of the distribution. We however note that since minimum wage legislation was not specifically targeting women, the trend of the wage gap at the 10th percentile suggests that an "unintended" outcome of the minimum wage legislation and by

[28] The analysis was carried out using stata codes from Fortin et al. (2011).

extension the Basic Conditions of employment Act number 75 of 1997, which allows the minister of labour to determine minimum wages for vulnerable sectors has been the narrowing of the wage gap at the bottom of the wage distribution.

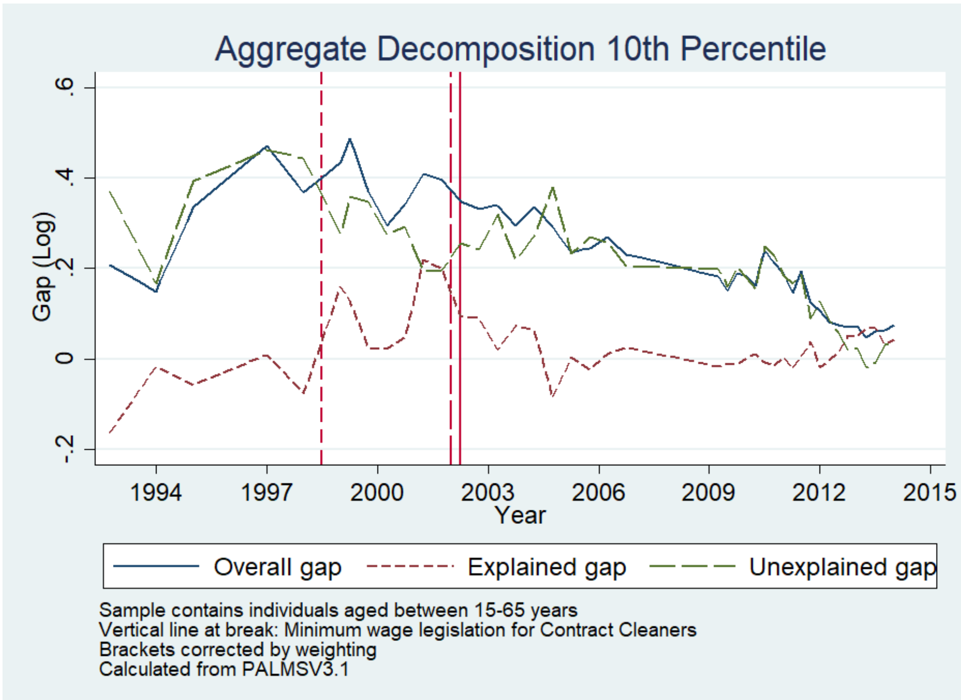


Figure 19: RIF decomposition: 10th Percentile

Notes: Omitted groups: African, single, with primary education and with less than 10 years of experience, from the Western Cape, in an elementary occupation in the manufacturing sector.

The trend at the median differs from what we see at the 10th percentile in that the wage gap does not seem to have changed much over time. However, like in the case of the mean and the 10th percentile, the overall gap seems to mimic the trend of the unexplained gap. At the median however, the unexplained gap does not seem to be declining over time meaning that the "discrimination" component is not declining. The expectation is that at the very least, anti-discrimination laws would have been more binding at the median. This is because the labour market policies such as the Labour Relations Act, number 66 of 1995, the Basic Conditions of Employment Act, number 75 of 1997, the Employment Equity Act number 55 of 1998 and the Black Economic Empowerment Act, number 53 of 2003 specifically targeted eliminating inequalities in the labour market and especially in formal employment in the public and private sectors where we are most likely to locate the median worker. The Employment Equity Act required employers to enforce affirmative action, while the Labour Relations Act secured the right of workers to unionise and the Skills Development Act compelled employers to extend training to previously disadvantaged groups including women.

The puzzle however is that the unexplained gap at the median is persistent suggesting that anti-discrimination laws have been less successful in reducing discrimination in the labour market. The median gap was at 0.301 log points in 1993, by 1999 it was at 0.24 log points and it has not moved much since, recording a figure of 0.21 log points in 2014. What is visible from these trends and the descriptive analysis however, is that the post-apartheid government has performed better at improving human capital skills for women. This can be inferred from the negative explained gap which suggests that if women's skills were at the same level as those of men, the gender wage gap in South Africa would be much wider.

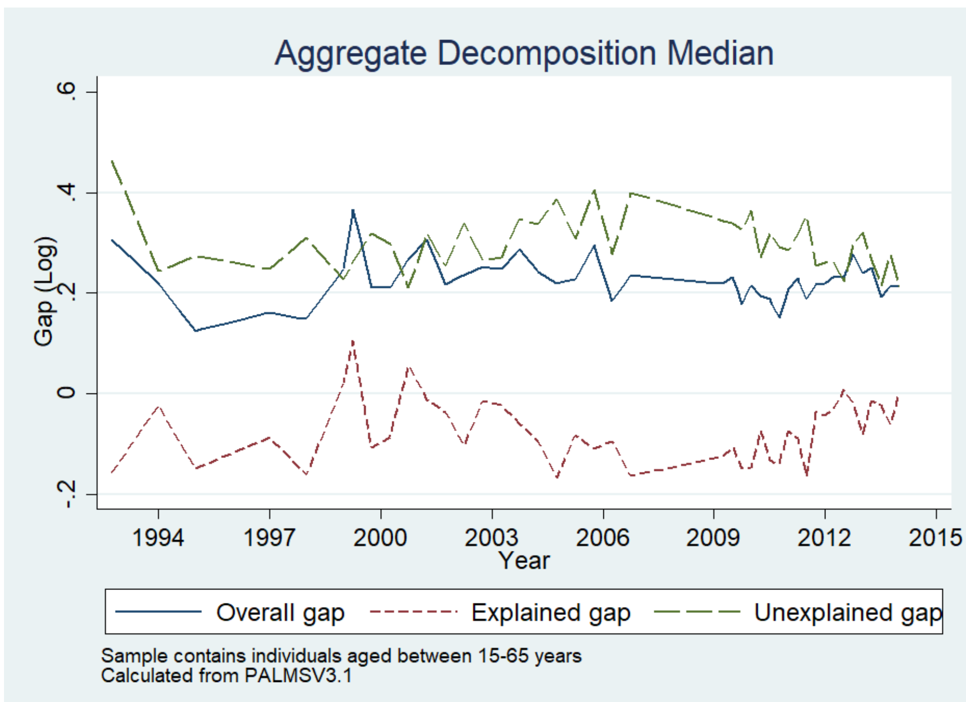


Figure 20: RIF decomposition at the median

Notes: Omitted groups: African, single, with primary education and with less than 10 years of experience, from the Western Cape, in an elementary occupation in the manufacturing sector.

Compared to the median, the raw gap at the 90th percentile shows a lot of fluctuations with a modest decline overall. The gap was at 0.41 log points in 1993 dropped to 0.12 log points in 1997, was at 0.15 log points in 2007 and at 0.18 log points in 2014. The drop of the gap suggests that highly skilled women benefited more from affirmative action in terms of accessing high paying occupations. However, the high unexplained gap which is always greater than the overall gap and which seems to be expanding after

2005, points to greater discrimination at the top of the distribution. The high and persistent unexplained gap and the negative explained gap points to the conclusion reached above that anti-discrimination legislation has been less successful in reducing gender discrimination in the labour market. It also points towards existence of a glass ceiling phenomena for women in the South African labour market. It is possible that the trend exhibited at the mean of a persistent wage gap is as a result of the persistent gender wage gap at the top half of the distribution.

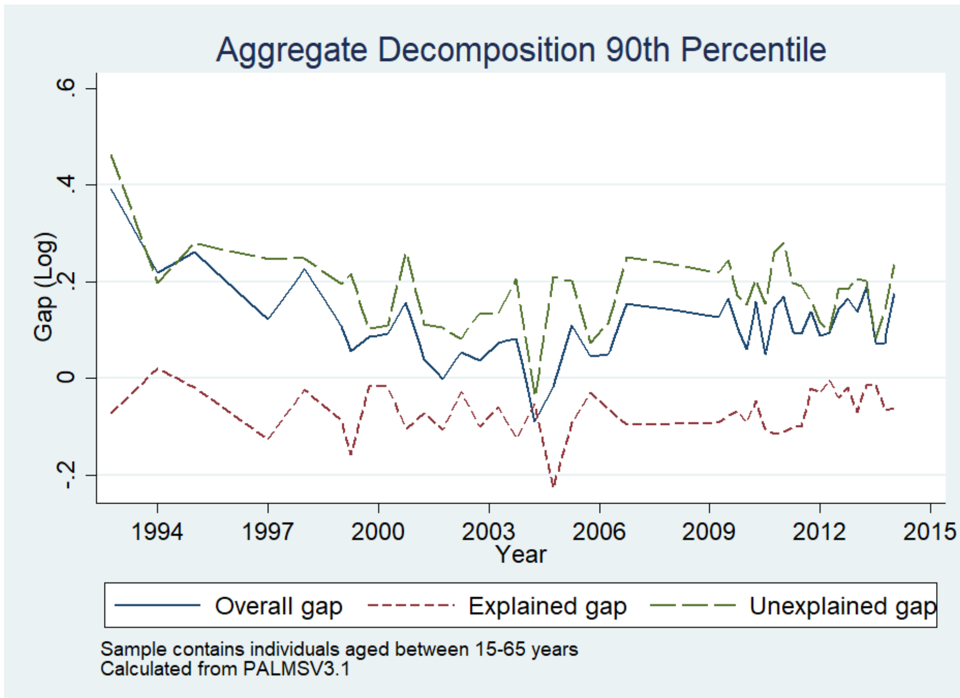


Figure 21: RIF decomposition at the 90th percentile

Notes: Omitted groups: African, single, with primary education and with less than 10 years of experience, from the Western Cape, in an elementary occupation in the manufacturing sector.

3.5.4.3 Detailed Decomposition

In addition to the aggregate decomposition we performed a detailed decomposition which shows the contribution of each explanatory variable to the explained and unexplained gap. A positive contribution of a variable to an explained gap means that men have an advantage in that variable that is, the $(\bar{X}_m - \bar{X}_f)'\hat{\beta}_f$ term is positive. A negative contribution means that women have an advantage in that variable that is, $(\bar{X}_m - \bar{X}_f)'\hat{\beta}_f$ is negative and therefore that variable contributes to narrowing the gender wage gap. A positive contribution of a variable to an unexplained gap means that men have an advantage in the rewards to that characteristic meaning that $\bar{X}_m'(\hat{\beta}_m - \hat{\beta}_f)$ is positive and therefore the variable contributes to the widening of the unexplained gap whereas a negative contribution means that women have better rewards for that characteristic and therefore the characteristic contributes to narrowing the unexplained gap.

Results from the detailed decomposition are presented in tables 10, 11, 12 and 13 however for purposes of a clearer discussion, detailed decomposition results from the LFS 2007:2 are presented in table 8. The choice of the LFS 2007:2 is arbitrary and the discussion extends to the complete results.

Across the entire wage distribution and in all waves, the explained gap is small and mostly negative except at the 10th percentile. Contributing positively to the explained gap across the wage distribution is industry of employment, union and marital status. Industry of employment is the most important characteristic in explaining the expansion of the wage gap across the wage distribution. Results from 2007 show that the contribution of industry to the gender wage gap declines monotonically across the wage distribution expanding the gap by 133.26% at the 10th percentile, by 52.56% at the median and by 19.71% at the 90th percentile. The contribution of the industry variable to the explained gap at the mean is 130.72%. The result is not surprising because at the bottom of the wage distribution women are concentrated in the low skill and low pay industries of domestic services, agriculture and retail trade. Although this effect is significant, results from the other waves show that it is declining over time.

The importance of unionisation is stable across the wage distribution increasing the gender gap by 45.73% at the 10th percentile and by 16.81% at the 90th percentile in 2007. This result is not surprising as female employees at the 10th percentile are less likely to be unionised due to the nature of the occupations they are employed in for example domestic work which happens in the private homes of employers. Marital status is important in contributing to the explained gap at the 90th percentile expanding the gap by 32.04% and only by 6.74% at the 10th percentile. At the mean marital status widens the gap by 18.43% coming from the fact that there are more married men on average in our sample.

Table 8: Detailed decomposition

Variable	10th		median		90th		mean	
	LFS07:2		LFS07:2		LFS07:2		LFS07:2	
Logw_Male	1.09(0.0234)		2.25(0.0348)		3.73(0.0692)		2.31(0.0349)	
Logw_Female	0.86(0.0277)		2.03(0.0456)		3.58(0.0495)		2.16(0.0351)	
Overall_gap	0.23(0.0362)		0.23(0.0574)		0.15(0.0851)		0.16(0.0495)	
Explained	0.0457(0.0319)		-0.13(0.0628)		-0.0827(0.0534)		-0.0586(0.0455)	
	[19.53]		[-56.58]		[-56.26]		[-37.56]	
Unexplained	0.19(0.0335)		0.36(0.0473)		0.23(0.0717)		0.21(0.0285)	
	[80.34]		[156.58]		[156.46]		[137.18]	
Covariates	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained	Explained	Unexplained
Experience	0.00232	-0.121	-0.00340	0.0165	0.00465	-0.118	-0.00173	-0.0158
	(0.00516)	(0.0809)	(0.00430)	(0.0796)	(0.00605)	(0.136)	(0.00418)	(0.0667)
	[5.08]	[-64.36]	[2.64]	[4.62]	[-5.62]	[-51.30]	[2.95]	[-7.38]
married	0.00308	0.0518	0.02	0.0306	0.03	0.0321	0.01	0.07
	(0.00594)	(0.0359)	(0.00760)	(0.0416)	(0.0117)	(0.0643)	(0.00555)	(0.0296)
	[6.74]	[27.55]	[-12.02]	[8.57]	[-32.04]	[13.96]	[-18.43]	[30.70]
province	0.00140	-0.11	-0.00579	0.28	-0.00305	0.0290	-0.00262	0.0359
	(0.0101)	(0.0604)	(0.00707)	(0.101)	(0.00551)	(0.198)	(0.00721)	(0.0669)
	[3.06]	[-57.98]	[4.49]	[77.03]	[3.69]	[12.61]	[4.47]	[16.78]
race	-0.00344	-0.00126	-0.0214	-0.00378	-0.00854	0.13	-0.0149	0.0140
	(0.00328)	(0.0126)	(0.0188)	(0.0208)	(0.00957)	(0.0501)	(0.0123)	(0.0159)
	[-7.53]	[-0.67]	[16.59]	[-1.06]	[10.33]	[57.83]	[25.43]	[6.54]
education	-0.04	-0.16	-0.06	0.00132	-0.06	0.115	-0.06	0.0308
	(0.0116)	(0.0884)	(0.0165)	(0.0740)	(0.0222)	(0.0756)	(0.0159)	(0.0447)
	[-91.68]	[-86.70]	[46.12]	[0.37]	[76.30]	[50.00]	[103.07]	[14.39]
occupation	0.0131	0.0462	-0.13	-0.0361	-0.06	0.111	-0.07	0.0198
	(0.0190)	(0.0699)	(0.0371)	(0.0694)	(0.0309)	(0.0700)	(0.0195)	(0.0385)
	[28.67]	[24.57]	[100.00]	[-10.11]	[75.09]	[48.26]	[116.21]	[9.25]
industry	0.06	-0.0560	0.07	-0.00265	0.0163	0.0639	0.08	-0.0840
	(0.0221)	(0.0632)	(0.0239)	(0.0947)	(0.0415)	(0.151)	(0.0179)	(0.0596)
	[133.26]	[-29.79]	[-52.56]	[-0.74]	[-19.71]	[27.78]	[-130.72]	[-39.25]
union	0.02	-0.0283	0.03	-0.0452	0.01	-0.00655	0.02	-0.0269
	(0.00622)	(0.0204)	(0.0107)	(0.0323)	(0.00811)	(0.0546)	(0.00615)	(0.0182)

	[45.73]	[-15.05]	[-27.05]	[-12.66]	[-16.81]	[-2.85]	[-35.15]	[-12.57]
public sector	-0.01	-0.00552	-0.03	0.0155	-0.00731	0.0157	-0.02	-0.0133
	(0.00432)	(0.0161)	(0.0101)	(0.0187)	(0.00725)	(0.0398)	(0.00676)	(0.0129)
	[-23.19]	[-2.94]	[22.02]	[4.34]	[8.84]	[6.83]	[32.25]	[-6.21]
Constant		0.57		0.106		-0.145		0.188
		(0.150)		(0.182)		(0.313)		(0.122)
		[305.32]		[29.69]		[-63.04]		[87.85]
Observations	15,388	15,388	15,388	15,388	15,388	15,388	15,388	15,388

Standard errors in parentheses and Percentages in squared brackets. Omitted category Single, African, from Western Cape, non-unionised, with primary school education or lower, with less than 10 years of experience, in an elementary occupation in the manufacturing sector in the public sector

The most important factors contributing negatively to the explained gap are public sector, occupation and education. The most important factor in narrowing the explained gap across the entire wage distribution is education. The contribution of education is stable across the wage distribution narrowing the gap by 91.68% at the 10th percentile and by 76.30% at the 90th percentile in 2007. At the mean this contribution is 103.07%. The result makes sense because from the summary statistics on average women have more years of schooling and especially there are more women than men with tertiary education in South Africa which contributes the highest negative effect. Looking at the entire series, the negative effect of education on the explained gap has been declining over time at the 10th percentile while it has been stable at the median and increasing over time at the 90th percentile.

The contribution of occupation is most important at the median reducing the explained gap by 100% in 2007. At the mean, occupation contributes negatively to the explained gap by 116.21%. This effect is mostly contributed to by a high negative effect from clerks, professional, technical and associate professionals. Employed women dominate the clerks, technical and associate professionals occupations with women constituting over 60% of the clerks. The negative effect is however tempered by a strong positive effect from occupations such as legislation and management, machine operators and crafts. This suggests that the occupational-barring referred to in Hinks (2002) and Rospabé (2001) is still strong in the Post-Apartheid Labour market. Public sector contributed to narrowing the explained gap by 22% at the median and by 23% at the 10th percentile and by only 8% at the 90th percentile. Descriptive statistics revealed that in the South African labour market there are more women than men employed in the public sector while OLS regressions reveal that employment in the public sector is more important for women than for men. Therefore, the fact that public sector contributes negatively to the explained gap is not surprising.

Contributing positively to the unexplained gap across the entire wage distribution is marital status and occupation. A positive contribution of marital status variable to the unexplained gap means that across the entire wage distribution, the returns to marriage are higher for men than for women. The difference in returns to marriage are however greatest at the 10th percentile increasing the unexplained gap by 27.55% in 2007. At the mean, marital status expanded the unexplained gap by 30.7%. OLS regression results showed that although at the mean married women earn higher than single women, they are still disadvantaged relative to men as men get higher returns to marriage than women (see table 9). This suggests that there is some preferential treatment in terms of pay for men. It could be the case that married men are viewed as more committed and likely to invest more time in the labour market whereas women are likely to invest less time as their loyalty is split between the labour market and the family. This suggests that the South African society is still patriarchal in nature where women are viewed as the main care givers and men the main 'bread winners'. It is also possible that the positive relationship between marriage and earnings among men is partly reflecting selection. This is because in South Africa, where marriage rates are low, higher earning men are more likely to marry (see Casale & Posel 2010).

Occupation contributes positively to the unexplained gap at the 10th percentile and at the 90th percentile but contributes negatively to the unexplained gap at the median. The negative contribution of occupation to the unexplained gap is due to women receiving better rewards in most occupations even those seemingly dominated by men such as management. We however apply caution in the interpretation of this result as it is not invariant to the choice of omitted category[29]. For the same reason caution is exercised in interpreting the contribution of education and industry of employment to the unexplained gap.

Education and industry of employment contribute negatively to the unexplained gap. Women have on average more educational qualifications than men and they receive better rewards for their educational

[29] In our case the omitted category is single, African, from Western Cape, not unionized, with primary school education or lower, with less than 10 years of experience, in an elementary occupation in the manufacturing sector in the public sector

characteristics and this has helped reduce the unexplained gap. This is however only true for the 10th percentile and median. Education contributes positively to the unexplained gap at the 90th percentile. This is because despite women having better educational qualifications men receive higher returns for their education at the top of the distribution causing an expansion of the unexplained gap. Across the wage distribution except at the 90th percentile, the contribution of industry to the unexplained component is negative but not significant at the 5 percent level. This result could be because women who are in male dominated industries such as finance and mining do receive better rewards than men. The problem is that they are too few for the effect to be significant.

3.6 DISCUSSION

In this chapter, we sought to re-examine the gender wage gap in the South African labour market. The inclusion of PSLSD 1993 gave us an alternative baseline period as the survey was carried out just before the installation of a new democratic government. The OHS 1994 and the OHS 1995 have been used before as baseline waves but data quality issues regarding the sampling design and inconsistencies in the measurement of domestic work necessitates the use of an alternative baseline if only for comparison purposes. Additionally, using a longer period helped us illuminate the breaks in the data and explain some peculiar trends of the gender wage gap raised in the literature.

Unsurprisingly, we find that the choice of baseline can determine the substantive conclusions drawn. For example, the findings that the gender wage gap had not declined in the period 1995 to 2004 (Ntuli 2007b; Grün 2004) were due to the use of OHS 1995 as the baseline year. Contrary to previous results in the literature, our analysis shows that there has been a decline of the gender wage gap at the mean from about 0.34 log points (about 40%) in 1993 to 0.15 log points (about 16%) in 2014. We attribute the decline at the mean to increased wages at the bottom of the distribution as a result of minimum wage legislation. Important to note, however, is that the gap declined until 2007 and has been stagnant since, oscillating at 16 %. We attribute this stagnation to the persistent gender wage gap at the top of the wage distribution.

We could not reproduce Winter’s (1999) result that women earned 87% of men’s wages in 1994. Our estimate of the gap in 1994 was 0.2 log points (approximately 22%) which means that according to our results women earned only 78% of men's wages this year. This difference in results is due to the inconsistency in the classification of domestic workers in this year and hence the importance of addressing data quality issues in any analysis.

In addition to these data quality issues, our study further finds several other substantive results not previously evident in the literature. Examination of the long run trend of the gender wage gap shows that there has been a substantial decline of the gender wage gap at the 10th percentile. We attribute most of this decline to wage structure effects arising from minimum wage legislation for low income groups. The effect of the minimum wage legislation for low income earners, including contract cleaners, domestic workers and agricultural workers, has been to raise the earnings of these groups (which are mostly comprised of women) and therefore leading to the narrowing of the gender wage gap at the bottom end of the wage distribution. At the 90th percentile, there was a decline in the gender gap between 1993 and 2005 but this trend has reversed exhibiting a continually increasing unexplained gap in recent years. This result is robust to different specifications, reference wage structure and decomposition method.

On the contrary, there has not been much change in the gender wage gap at the median. This is surprising given the efforts by the South African government to reduce inequality in the labour market by implementing several anti-discrimination policies that specifically targeted women. This result is not however unique for South Africa. Polachek (2014) reports that despite the fact that equal pay legislation was introduced in the United States in 1964, the gender wage gap persists at about 22%. Similarly, for both the United Kingdom and France where Equal Pay legislation was enacted in 1970 and 1972 respectively, the gender wage gap in these countries remains at about 21 % and 17% respectively (Polachek 2014).

Descriptive analysis of the data showed that employed women in South Africa have on average an advantage in years of schooling. Additionally, employed women are less likely to drop out of high school, and more likely to have some tertiary education and just as likely or more likely to have a university degree. This has led to an increasing number of women in high skilled occupations, such as professionals, technical, associate professionals and clerks. The human capital theory stipulates that regardless of gender, the group with better human capital characteristics will have a wage advantage. That women receive lower rewards than men for their human capital characteristics has led to an increasingly negative explained gap and an expanding unexplained gap (usually associated with discrimination), especially at the 90th percentile. What the trends at the median and 90th percentile therefore tell us, is that the post-apartheid government has been successful in improving the human capital characteristics of women but less so in reducing the level of wage inequality, especially at the median and top of the wage distribution.

Like previous studies in the literature (Bhorat & Goga 2013; Ntuli 2007b) we found that the wage gap in the South African labour market exhibits a 'Sticky floor effect' with the gender wage gap being highest at the 10th percentile. However, we add to the evidence base by showing that since 2007, South Africa has also exhibited a 'glass ceiling effect' (the 90th percentile gap exceeds the 10th percentile gap) due to the large decline of the 10th percentile gap over time and the expansion of the gender wage gap at the top.

Finally, the detailed decompositions reveal that education is an important factor in increasing the unexplained gap at the 90th percentile but contributes to reducing it in other parts of the wage distribution. Occupation of employment also contributes positively to the unexplained gap at the 90th percentile. This means that there is a type of a ceiling for highly qualified women and adds to the findings of Rospabé (2001) and Hinks (2002) of occupational barring which may be the reason for the persistent gender wage gap at the top of the distribution. Although over time there has been an increase of women in management, women are still under-represented at this level and there are occupations that are still male dominated such as crafts and machine operators. Industry of employment is an important factor in reducing the unexplained gap although, this result is not significant. We suspect the reason for this is that there are very few women in the most lucrative industries such as mining.

A limitation of our study is that it may suffer from omitted variable bias. O'Neill & O'Neill (2006) finds that differences in schooling can no longer explain the differences in pay between men and women in the United States and that factors that affect choices made by men and women in the time and energy devoted to careers are more useful when trying to explain the gender wage gap. This finding is supported by Fortin (2008) who found that non-cognitive factors such as the importance men and women place on money and work and the importance of people or family affect the gender wage gap in a small but non-trivial way among young adults in America. This is because such factors affect the choices men and women make regarding the time, effort and responsibility allocated between labour market work and home. The fact that the explained gap is negative could just mean that we are not looking at important variables that affect wages. Future research should therefore focus on factors that determine time and effort allocation between the home and the labour market.

In sum, there has been progress in the reduction of the wage gap at the bottom of the wage distribution which we link to the introduction of minimum wage laws for low earning sectors. Affirmative action policies have however not been as successful in reducing the gender wage gap as shown by the trend in the median gap. The gender wage gap persists and given the “unintended” effect of the minimum wage legislation, we suggest that policy should be pointed towards improving the wages of women in all parts of the wage distribution. This requires removing barriers in the highest paying occupations such as management and increasing the number of women in male dominated industries. The question therefore is why there are so few women in mining, crafts, finance and machine operation? This question is tough to answer as it requires examining social and cultural norms and pre-labour market discrimination.

Cross sectional data is by its nature a snap shot picture of the labour market as it collects labour market information at a point in time. A cross sectional analysis like the one carried out in this chapter therefore compares individuals from different generations and therefore may not capture factors such as changes in social norms that may affect the gender wage gap. Chapter four of this thesis constructs cohort data from repeated cross sections and asks whether recent cohorts of women have a different labour market experience from those women born several decades earlier. It is possible that by following cohorts of women through their lives, we may find additional declines in the gender wage gap over time.

Table 9: Results from the OLS regression separately by gender

Table 9: Results from the OLS regression separately by gender

VARIABLES	OHS94		OHS99		LFS03:2		LFS07:2		QLFS10:4		QLFS13:4	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Less20yrs	0.171***	0.283***	0.315***	0.322***	0.297***	0.344***	0.152***	0.178**	0.108***	0.162***	0.102**	0.107**
	(0.0389)	(0.0386)	(0.0505)	(0.0542)	(0.0364)	(0.0435)	(0.0527)	(0.0737)	(0.0306)	(0.0311)	(0.0432)	(0.0433)
Less30yrs	0.371***	0.343***	0.510***	0.529***	0.548***	0.521***	0.355***	0.328***	0.193***	0.292***	0.185***	0.155***
	(0.0403)	(0.0431)	(0.0605)	(0.0560)	(0.0404)	(0.0438)	(0.0631)	(0.0943)	(0.0354)	(0.0333)	(0.0506)	(0.0457)
more30yrs	0.351***	0.351***	0.559***	0.598***	0.655***	0.627***	0.476***	0.470***	0.322***	0.362***	0.289***	0.235***
	(0.0447)	(0.0454)	(0.0620)	(0.0585)	(0.0433)	(0.0451)	(0.0632)	(0.0815)	(0.0383)	(0.0366)	(0.0560)	(0.0473)
Eastern Cape	-0.0839**	-0.174***	-0.410***	-0.487***	-0.259***	-0.202***	-0.247***	-0.149*	-0.134***	-0.132***	-0.236***	-0.212***
	(0.0408)	(0.0441)	(0.0762)	(0.0744)	(0.0488)	(0.0568)	(0.0574)	(0.0778)	(0.0438)	(0.0437)	(0.0665)	(0.0591)
Northern Cape	-0.466***	-0.527***	-0.194***	-0.316***	-0.0411	-0.299***	-0.0476	-0.182***	-0.209***	-0.0200	-0.0333	-0.107
	(0.0378)	(0.0501)	(0.0659)	(0.0791)	(0.0559)	(0.0586)	(0.0610)	(0.0653)	(0.0427)	(0.0481)	(0.0684)	(0.0738)
Free State	-0.672***	-0.615***	-0.367***	-0.569***	-0.196***	-0.263***	-0.0745	-0.155*	-0.253***	-0.205***	-0.219***	-0.283***
	(0.0532)	(0.0564)	(0.0599)	(0.0759)	(0.0467)	(0.0611)	(0.0696)	(0.0821)	(0.0465)	(0.0464)	(0.0637)	(0.0579)
KwaZulu-Natal	0.0120	-0.0454	-0.0818	-0.204***	-0.00847	-0.180***	-0.0574	-0.108	-0.0634	-0.118***	-0.102*	-0.282***
	(0.0404)	(0.0445)	(0.0620)	(0.0665)	(0.0437)	(0.0533)	(0.0604)	(0.0802)	(0.0437)	(0.0439)	(0.0614)	(0.0567)
North Western	-0.0893*	-0.102*	-0.0413	-0.157**	0.0511	-0.166***	-0.0219	-0.107	0.0647	-0.0328	0.136**	-0.131**
	(0.0486)	(0.0591)	(0.0627)	(0.0754)	(0.0453)	(0.0600)	(0.0579)	(0.0805)	(0.0480)	(0.0520)	(0.0659)	(0.0595)
Gauteng	0.228***	0.243***	0.129**	0.0821	0.218***	0.182***	0.120**	0.132	0.0671*	0.0784*	0.0616	0.00260
	(0.0386)	(0.0401)	(0.0553)	(0.0641)	(0.0417)	(0.0515)	(0.0547)	(0.0976)	(0.0391)	(0.0408)	(0.0585)	(0.0505)
Mpumalanga	-0.217***	-0.257***	-0.116**	-0.264***	0.0157	-0.0962*	0.0106	-0.146*	-0.00795	-0.0324	0.00867	-0.172***
	(0.0557)	(0.0765)	(0.0576)	(0.0726)	(0.0478)	(0.0580)	(0.0756)	(0.0847)	(0.0489)	(0.0496)	(0.0711)	(0.0606)
Limpopo	-0.134**	-0.215***	-0.165**	-0.226***	-0.0911	-0.263***	-0.142**	-0.276***	-0.251***	-0.247***	-0.261***	-0.428***
	(0.0546)	(0.0693)	(0.0652)	(0.0729)	(0.0559)	(0.0636)	(0.0635)	(0.0868)	(0.0537)	(0.0529)	(0.0687)	(0.0626)
Coloured	0.138***	0.152***	0.0848	0.217***	0.253***	0.371***	0.245***	0.376***	0.175***	0.251***	0.0447	0.104**
	(0.0301)	(0.0338)	(0.0558)	(0.0577)	(0.0415)	(0.0491)	(0.0498)	(0.0917)	(0.0358)	(0.0387)	(0.0568)	(0.0506)
White	0.498***	0.443***	0.668***	0.523***	0.832***	0.727***	0.811***	0.657***	0.788***	0.644***	0.538***	0.598***
	(0.0349)	(0.0400)	(0.0595)	(0.0670)	(0.0451)	(0.0499)	(0.0788)	(0.121)	(0.0372)	(0.0351)	(0.0568)	(0.0505)
Indian	0.228***	0.248***	0.299***	0.565***	0.548***	0.667***	0.489***	0.585***	0.545***	0.725***	0.315***	0.449***
	(0.0434)	(0.0567)	(0.0878)	(0.0973)	(0.0589)	(0.0664)	(0.0721)	(0.0979)	(0.0656)	(0.0589)	(0.0907)	(0.0917)
<Matric	0.719***	0.701***	0.572***	0.761***	0.564***	0.578***	0.426***	0.515***	0.349***	0.343***	0.276***	0.276***
	(0.0282)	(0.0346)	(0.0382)	(0.0427)	(0.0260)	(0.0323)	(0.0384)	(0.0505)	(0.0299)	(0.0305)	(0.0387)	(0.0349)
Matric	1.140***	1.253***	0.967***	1.463***	1.101***	1.257***	0.884***	1.090***	0.787***	0.913***	0.671***	0.771***
	(0.0394)	(0.0418)	(0.0480)	(0.0550)	(0.0337)	(0.0417)	(0.0514)	(0.0719)	(0.0341)	(0.0360)	(0.0461)	(0.0404)
Tertiary	1.578***	1.703***	1.680***	2.259***	1.892***	2.159***	1.796***	1.979***	1.581***	1.825***	1.506***	1.601***
	(0.0533)	(0.0496)	(0.0691)	(0.0521)	(0.0465)	(0.0397)	(0.0543)	(0.0650)	(0.0427)	(0.0395)	(0.0565)	(0.0457)
Married	0.130***	-0.0485*	0.250***	0.0719**	0.268***	0.0710***	0.280***	0.127***	0.199***	0.0449**	0.202***	0.0457
	(0.0291)	(0.0269)	(0.0358)	(0.0335)	(0.0253)	(0.0269)	(0.0351)	(0.0451)	(0.0240)	(0.0218)	(0.0337)	(0.0278)
Constant	0.998***	0.839***	1.015***	0.589***	0.714***	0.516***	1.139***	0.866***	1.327***	1.025***	1.388***	1.325***
	(0.0472)	(0.0518)	(0.0678)	(0.0736)	(0.0499)	(0.0604)	(0.0638)	(0.107)	(0.0505)	(0.0509)	(0.0700)	(0.0651)
Observations	8,046	5,318	6,355	4,806	7,707	5,772	8,767	6,855	8,486	7,738	8,373	7,886
R-squared	0.472	0.504	0.354	0.490	0.506	0.565	0.501	0.520	0.440	0.495	0.262	0.332

Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Omitted category Single, African, from western cape, with primary school education or lower

Table 10: Mean Decomposition

Table 10: Mean Decomposition

VARs	PSLSD 1993		OHS 1997		OHS 1999		LFS03:2		LFS07:2		QLFS10:4		QLFS14:4	
Male	2.416***(0.0234)		2.254***(0.0152)		2.124***(0.0196)		2.129***(0.0164)		2.314***(0.0349)		2.354***(0.0145)		2.229***(0.0179)	
Female	2.073***(0.0264)		2.004***(0.0179)		1.863***(0.0248)		1.945***(0.0219)		2.159***(0.0351)		2.198***(0.0157)		2.079***(0.0171)	
Difference	0.343***(0.0353)		0.251***(0.0235)		0.261***(0.0316)		0.184***(0.0273)		0.156***(0.0495)		0.156***(0.0214)		0.150***(0.0247)	
Explained	-0.144***(0.0438)		-0.0359(0.0259)		0.0355(0.0401)		-0.0419(0.0349)		-0.0586(0.0455)		-0.0877***(0.0216)		-0.00501(0.0240)	
	0.487***(0.0405)		0.287***(0.0239)		0.225***(0.0367)		0.226***(0.0293)		0.214***(0.0285)		0.243***(0.0198)		0.156***(0.0268)	
Covariates	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl
	0.00763**	0.115**	0.00985***	0.0545	0.00658**	0.0238	-8.98e-05	-0.00102	-0.00173	-0.0158	-0.00439**	-0.0584*	-0.00534***	-0.00645
	(0.00343)	(0.0547)	(0.00247)	(0.0486)	(0.00326)	(0.0597)	(0.00290)	(0.0413)	(0.00418)	(0.0667)	(0.00187)	(0.0330)	(0.00198)	(0.0471)
Married	-0.00476*	0.0888***	8.04e-05	0.0709***	0.00540	0.0790***	0.00223	0.100***	0.0108*	0.0657**	0.00212	0.0535***	0.00793***	0.0343
	(0.00265)	(0.0289)	(0.00339)	(0.0229)	(0.00473)	(0.0304)	(0.00361)	(0.0201)	(0.00555)	(0.0296)	(0.00182)	(0.0172)	(0.00277)	(0.0237)
Province	-0.0177***	-0.00633	-0.00128	0.0516	-0.000804	0.0726	0.000648	0.128***	-0.00262	0.0359	0.00219	0.0326	0.00602**	0.101*
	(0.00660)	(0.0614)	(0.00297)	(0.0471)	(0.00636)	(0.0595)	(0.00493)	(0.0447)	(0.00721)	(0.0669)	(0.00244)	(0.0429)	(0.00260)	(0.0614)
Race	-0.0305***	0.0475**	-0.0118**	0.0341**	-0.0145***	0.0181	-0.0199***	0.0216**	-0.0149	0.0140	-0.0176***	0.0134	-0.000787	-0.0194
	(0.00824)	(0.0191)	(0.00534)	(0.0134)	(0.00549)	(0.0150)	(0.00612)	(0.0104)	(0.0123)	(0.0159)	(0.00478)	(0.0116)	(0.00341)	(0.0168)
Education	-0.0156**	0.105***	-0.0763***	-0.0706**	-0.0621***	-0.0592*	-0.0519***	0.0531*	-0.0604***	0.0308	-0.0555***	-0.0633*	-0.0305***	-0.0909*
	(0.00632)	(0.0340)	(0.00849)	(0.0297)	(0.0101)	(0.0348)	(0.00770)	(0.0281)	(0.0159)	(0.0447)	(0.00802)	(0.0382)	(0.00556)	(0.0516)
	-0.0329	-0.209***	-0.0316**	0.0364	-0.0470*	0.0241	-0.0655***	0.0164	-0.0681***	0.0198	-0.0285*	0.0703**	-0.0129	-0.0307
	(0.0271)	(0.0518)	(0.0135)	(0.0339)	(0.0248)	(0.0505)	(0.0200)	(0.0321)	(0.0195)	(0.0385)	(0.0151)	(0.0322)	(0.0194)	(0.0404)
Industry	-0.0708***	0.118**	0.111***	-0.00832	0.152***	-0.0992	0.0921***	-0.0570	0.0766***	-0.0840	0.0114	-0.113**	0.0332***	-0.131**
	(0.0259)	(0.0550)	(0.0185)	(0.0400)	(0.0295)	(0.0766)	(0.0184)	(0.0505)	(0.0179)	(0.0596)	(0.00985)	(0.0475)	(0.0121)	(0.0656)
Union	0.0250***	-0.00599	0.0199***	-0.0443***	0.0416***	-0.0726***	0.0339***	-0.0255*	0.0206***	-0.0269	0.0128***	-0.00964	0.00984***	-0.00452
	(0.00524)	(0.0168)	(0.00328)	(0.0140)	(0.00691)	(0.0255)	(0.00478)	(0.0144)	(0.00615)	(0.0182)	(0.00289)	(0.0132)	(0.00260)	(0.0173)
	-0.00475	-0.0201	-0.0559***	-0.0450***	-0.0455***	-0.0335*	-0.0335***	-0.0246**	-0.0189***	-0.0133	-0.0102***	0.00950	-0.0124***	0.0131
	(0.00412)	(0.0158)	(0.00666)	(0.0124)	(0.00906)	(0.0182)	(0.00646)	(0.0109)	(0.00676)	(0.0129)	(0.00250)	(0.00904)	(0.00320)	(0.0118)
Constant		0.254**		0.207**		0.272**		0.0144		0.188		0.308***		0.290**
		(0.116)		(0.0899)		(0.131)		(0.0896)		(0.122)		(0.0854)		(0.127)
	6,294	6,294	15,383	15,383	10,737	10,737	13,314	13,314	15,388	15,388	15,825	15,825	15,003	15,003

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Omitted category Single, African, from western cape, non unionised, with primary school education or lower, in an elementary occupation in the manufacturing sector in the public sector

Table 11: RIF Decomposition 10th Percentile

Table 11: RIF-Decomposition: 10th percentile

VARs	PSLSD 1993		OHS 1997		OHS 1999		LFS03:2		LFS07:2		QLFS10:4		QLFS14:4	
Male	0.751***(0.0526)		0.877***(0.0202)		0.689***(0.0339)		0.818***(0.0254)		1.093***(0.0234)		1.118***(0.0174)		0.836***(0.0275)	
Female	0.544***(0.0532)		0.400***(0.0303)		0.250***(0.0475)		0.490***(0.0318)		0.860***(0.0277)		0.937***(0.0171)		0.764***(0.0215)	
Difference	0.207***(0.0748)		0.478***(0.0364)		0.440***(0.0583)		0.328***(0.0406)		0.234***(0.0362)		0.181***(0.0244)		0.0721**(0.0349)	
Explained	-0.239***(0.0731)		0.0809**(0.0412)		0.270***(0.0772)		0.135**(0.0528)		0.0457(0.0319)		0.0130(0.0238)		0.0408(0.0316)	
	0.446***(0.0906)		0.397***(0.0445)		0.170**(0.0808)		0.193***(0.0593)		0.188***(0.0335)		0.168***(0.0315)		0.0313(0.0459)	
Covariates	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl
	0.0104	-0.0298	0.0185***	-0.0708	0.0118*	-0.123	-0.00211	-0.115	0.00232	-0.121	-0.00532**	-0.0707	-0.000356	-0.111
	(0.00682)	(0.167)	(0.00511)	(0.0783)	(0.00609)	(0.136)	(0.00498)	(0.0969)	(0.00516)	(0.0809)	(0.00267)	(0.0551)	(0.00258)	(0.0832)
Married	-0.0230**	0.00729	-0.0204**	0.156***	-0.0253*	0.263***	-0.00537	0.169***	0.00308	0.0518	-0.00185	0.0540*	0.00250	0.0904**
	(0.00895)	(0.0891)	(0.00844)	(0.0492)	(0.0134)	(0.0759)	(0.00943)	(0.0516)	(0.00594)	(0.0359)	(0.00296)	(0.0285)	(0.00435)	(0.0430)
Province	-0.0308*	-0.289**	0.00687	0.100	-0.000256	-0.00607	0.0124	0.00217	0.00140	-0.109*	0.00148	0.0430	0.00631**	0.0847
	(0.0158)	(0.122)	(0.00831)	(0.0745)	(0.0181)	(0.144)	(0.0124)	(0.0817)	(0.0101)	(0.0604)	(0.00448)	(0.0500)	(0.00247)	(0.0925)
Race	-0.00835	-0.00924	-0.00348	-0.00401	-0.00707	-0.0265	-0.0106***	0.00395	-0.00344	-0.00126	-0.00793***	-0.00820	0.00121	-0.0504**
	(0.00584)	(0.0377)	(0.00315)	(0.0205)	(0.00468)	(0.0298)	(0.00369)	(0.0172)	(0.00328)	(0.0126)	(0.00252)	(0.0132)	(0.00132)	(0.0246)
Education	-0.0254**	-0.0376	-0.0832***	-0.372***	-0.0828***	-0.204**	-0.0460***	-0.122	-0.0419***	-0.163*	-0.0273***	-0.0973	-0.0128**	-0.0434
	(0.0104)	(0.0862)	(0.0126)	(0.0699)	(0.0169)	(0.0997)	(0.0104)	(0.0866)	(0.0116)	(0.0884)	(0.00589)	(0.0838)	(0.00509)	(0.110)
	-0.0141	-0.144	-0.00825	0.146**	0.163***	0.0880	-0.0154	0.147*	0.0131	0.0462	0.00978	0.137**	-0.00303	0.0867
	(0.0364)	(0.132)	(0.0248)	(0.0719)	(0.0520)	(0.130)	(0.0399)	(0.0863)	(0.0190)	(0.0699)	(0.0224)	(0.0591)	(0.0301)	(0.0791)
Industry	-0.150**	0.193*	0.229***	0.0811	0.223***	-0.179	0.183***	-0.162*	0.0609***	-0.0560	0.0405***	-0.236***	0.0574***	0.00615
	(0.0639)	(0.114)	(0.0360)	(0.0725)	(0.0623)	(0.143)	(0.0375)	(0.0871)	(0.0221)	(0.0632)	(0.0148)	(0.0678)	(0.0218)	(0.113)
Union	0.00818	0.0455	0.0351***	-0.101***	0.0540***	0.0139	0.0462***	-0.0346	0.0209***	-0.0283	0.00844***	0.0210	-0.00255	0.0162
	(0.00872)	(0.0304)	(0.00574)	(0.0221)	(0.0118)	(0.0479)	(0.00719)	(0.0248)	(0.00622)	(0.0204)	(0.00229)	(0.0155)	(0.00286)	(0.0282)
	-0.00631	0.00146	-0.0932***	-0.0875***	-0.0668***	-0.0400	-0.0275***	0.0431**	-0.0106**	-0.00552	-0.00486	0.0176	-0.00791	-0.0298
	(0.00565)	(0.0374)	(0.0123)	(0.0248)	(0.0151)	(0.0369)	(0.00678)	(0.0213)	(0.00432)	(0.0161)	(0.00304)	(0.0128)	(0.00495)	(0.0209)
Constant		0.709**		0.549***		0.382		0.262		0.574***		0.307**		-0.0180
		(0.278)		(0.167)		(0.291)		(0.188)		(0.150)		(0.150)		(0.223)
	6,294	6,294	15,383	15,383	10,737	10,737	13,314	13,314	15,388	15,388	15,825	15,825	15,003	15,003

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Omitted category Single, African, from western cape, non unionised, with primary school education or lower, in an elementary occupation in the manufacturing sector in the public sector

Table 12: RIF-Decomposition: Median

VARs	pslsd93		ohs97		ohs99		LFS03:2		LFS07:2		QLFS10:4		QLFS14:4	
Male	2.441***		2.285***		2.118***		2.098***		2.253***		2.265***		2.129***	
	(0.0264)		(0.0185)		(0.0220)		(0.0193)		(0.0348)		(0.0186)		(0.0190)	
Female	2.140***		2.125***		1.875***		1.849***		2.025***		2.053***		1.919***	
	(0.0367)		(0.0255)		(0.0277)		(0.0247)		(0.0456)		(0.0230)		(0.0214)	
difference	0.301***		0.159***		0.243***		0.249***		0.228***		0.212***		0.211***	
	(0.0452)		(0.0315)		(0.0354)		(0.0314)		(0.0574)		(0.0296)		(0.0286)	
explained	-0.136**		-0.0262		0.0907*		0.0413		-0.129**		-0.117***		0.0274	
	(0.0595)		(0.0401)		(0.0503)		(0.0458)		(0.0628)		(0.0330)		(0.0332)	
unexplained	0.437***		0.186***		0.152***		0.208***		0.357***		0.329***		0.184***	
	(0.0586)		(0.0402)		(0.0505)		(0.0447)		(0.0473)		(0.0341)		(0.0377)	
Covariates	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl
Experience	0.00906**	0.0252	0.0123***	-0.0302	0.00491	0.0791	-0.00190	0.0290	-0.00340	0.0165	-0.00694***	-0.0551	-0.00762***	0.0613
	(0.00404)	(0.0781)	(0.00359)	(0.0641)	(0.00301)	(0.0693)	(0.00311)	(0.0550)	(0.00430)	(0.0796)	(0.00256)	(0.0527)	(0.00241)	(0.0568)
married	0.00178	0.105**	0.000214	0.0837**	0.0173***	0.0588	0.00780	0.0421	0.0155**	0.0306	0.00106	0.0863***	0.00854**	0.0112
	(0.00402)	(0.0446)	(0.00552)	(0.0359)	(0.00596)	(0.0371)	(0.00502)	(0.0288)	(0.00760)	(0.0416)	(0.00305)	(0.0276)	(0.00379)	(0.0291)
province	-0.0170**	0.0595	-0.00172	0.0323	-0.00101	0.0908	-0.00318	0.211***	-0.00579	0.275***	0.00287	0.0754	0.00631	0.230***
	(0.00762)	(0.0861)	(0.00393)	(0.0696)	(0.00487)	(0.0878)	(0.00463)	(0.0675)	(0.00707)	(0.101)	(0.00340)	(0.0678)	(0.00409)	(0.0678)
race	-0.0335***	-0.0236	-0.0202***	0.00518	-0.0168***	-0.00741	-0.0189***	0.0237*	-0.0214	-0.00378	-0.0253***	-0.0227	0.000519	-0.00157
	(0.00895)	(0.0275)	(0.00750)	(0.0195)	(0.00644)	(0.0183)	(0.00670)	(0.0143)	(0.0188)	(0.0208)	(0.00703)	(0.0175)	(0.00320)	(0.0187)
education	-0.0170**	0.105**	-0.0822***	-0.0154	-0.0561***	-0.0528	-0.0440***	0.0802*	-0.0595***	0.00132	-0.0538***	-0.125**	-0.0344***	-0.225***
	(0.00691)	(0.0517)	(0.0104)	(0.0462)	(0.0101)	(0.0462)	(0.00774)	(0.0435)	(0.0165)	(0.0740)	(0.00865)	(0.0629)	(0.00691)	(0.0685)
occupation	-0.0647	-0.401***	-0.0417	-0.0780	-0.0112	-0.0542	-0.0516*	-0.0978*	-0.129***	-0.0361	-0.0703***	-0.00475	0.00547	-0.0467
	(0.0407)	(0.0804)	(0.0264)	(0.0616)	(0.0332)	(0.0693)	(0.0297)	(0.0548)	(0.0371)	(0.0694)	(0.0250)	(0.0542)	(0.0283)	(0.0581)
industry	-0.0562	0.100	0.149***	-0.00414	0.164***	-0.151	0.140***	-0.0779	0.0678***	-0.00265	0.0290*	-0.104	0.0391**	-0.113
	(0.0364)	(0.101)	(0.0289)	(0.0737)	(0.0371)	(0.0929)	(0.0338)	(0.0698)	(0.0239)	(0.0947)	(0.0164)	(0.0726)	(0.0171)	(0.0824)
union	0.0482***	-0.00556	0.0298***	-0.0527**	0.0465***	-0.0470	0.0523***	-0.0429*	0.0349***	-0.0452	0.0230***	-0.0396*	0.0194***	0.00227
	(0.00970)	(0.0293)	(0.00515)	(0.0228)	(0.00771)	(0.0298)	(0.00735)	(0.0226)	(0.0107)	(0.0323)	(0.00515)	(0.0218)	(0.00424)	(0.0208)
public sector	-0.00631	-0.0325	-0.0717***	-0.0603***	-0.0565***	-0.0445*	-0.0391***	-0.0160	-0.0284***	0.0155	-0.0162***	0.00195	-0.01000***	0.0147
	(0.00552)	(0.0242)	(0.0102)	(0.0209)	(0.0113)	(0.0230)	(0.00780)	(0.0156)	(0.0101)	(0.0187)	(0.00393)	(0.0132)	(0.00386)	(0.0135)
Constant		0.505***		0.305**		0.281*		0.0564		0.106		0.517***		0.250
		(0.187)		(0.145)		(0.170)		(0.130)		(0.182)		(0.135)		(0.153)
Observations	6,294	6,294	15,383	15,383	10,737	10,737	13,314	13,314	15,388	15,388	15,825	15,825	15,003	15,003

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Omitted category Single, African, from western cape, non unionized, with primary school education or lower, in an elementary occupation in the manufacturing sector in the public sector

Table 13: RIF-Decomposition-90th Percentile

VARs	pslsd93		ohs97		ohs99		LFS03:2		LFS07:2		QLFS10:4		QLFS14:4	
Male	3.945***		3.541***		3.530***		3.535***		3.726***		3.818***		3.851***	
	(0.0414)		(0.0290)		(0.0339)		(0.0296)		(0.0692)		(0.0283)		(0.0317)	
Female	3.531***		3.420***		3.426***		3.505***		3.578***		3.750***		3.666***	
	(0.0318)		(0.0267)		(0.0336)		(0.0305)		(0.0495)		(0.0257)		(0.0251)	
difference	0.414***		0.121***		0.104**		0.0305		0.147*		0.0674*		0.185***	
	(0.0522)		(0.0394)		(0.0478)		(0.0425)		(0.0851)		(0.0382)		(0.0404)	
explained	-0.0586		-0.113***		-0.0675		-0.0957		-0.0827		-0.0801***		-0.0592**	
	(0.0528)		(0.0411)		(0.0727)		(0.0636)		(0.0534)		(0.0298)		(0.0291)	
unexplained	0.473***		0.234***		0.171**		0.126*		0.230***		0.148***		0.244***	
	(0.0643)		(0.0497)		(0.0808)		(0.0678)		(0.0717)		(0.0401)		(0.0409)	
Covariates	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl	Expl	Unexpl
Experience	0.0129**	0.340**	0.00438	0.227**	0.00839*	-0.175	0.00177	0.0908	0.00465	-0.118	-0.00710**	-0.0409	-0.00750**	-0.0337
	(0.00545)	(0.137)	(0.00408)	(0.115)	(0.00448)	(0.147)	(0.00298)	(0.0990)	(0.00605)	(0.136)	(0.00331)	(0.0778)	(0.00328)	(0.0896)
married	-0.00196	0.140**	0.00567	0.0437	0.000461	0.142**	-0.00518	0.161***	0.0265**	0.0321	0.00703*	-0.00713	0.00546	-0.0133
	(0.00434)	(0.0572)	(0.00705)	(0.0496)	(0.00958)	(0.0631)	(0.00789)	(0.0469)	(0.0117)	(0.0643)	(0.00396)	(0.0387)	(0.00463)	(0.0417)
province	-0.00768	0.0446	-0.00415	0.00859	0.00475	0.0365	0.00375	0.0542	-0.00305	0.0290	0.00235	-0.130	0.00444	0.0491
	(0.00736)	(0.170)	(0.00420)	(0.122)	(0.00582)	(0.134)	(0.00458)	(0.115)	(0.00551)	(0.198)	(0.00261)	(0.109)	(0.00274)	(0.103)
race	-0.0454***	0.176***	-0.0136*	0.0612*	-0.0116*	0.0781**	-0.0268***	0.0388	-0.00854	0.133***	-0.0169***	0.0697**	-0.00379	-0.00561
	(0.0135)	(0.0501)	(0.00739)	(0.0347)	(0.00609)	(0.0342)	(0.00819)	(0.0285)	(0.00957)	(0.0501)	(0.00534)	(0.0317)	(0.00533)	(0.0322)
education	-0.00893	0.152**	-0.0667***	0.0647	-0.0498***	0.00960	-0.0461***	0.181***	-0.0631***	0.115	-0.0690***	0.0340	-0.0290***	-0.0265
	(0.00814)	(0.0764)	(0.0123)	(0.0520)	(0.0129)	(0.0593)	(0.0111)	(0.0481)	(0.0222)	(0.0756)	(0.0107)	(0.0578)	(0.00675)	(0.0738)
occupation	0.0111	0.0128	-0.0587**	0.123**	-0.0974**	0.0625	-0.00959	-0.0307	-0.0621**	0.111	0.0243	0.0215	0.0151	-0.0246
	(0.0356)	(0.0854)	(0.0244)	(0.0532)	(0.0496)	(0.0782)	(0.0359)	(0.0538)	(0.0309)	(0.0700)	(0.0219)	(0.0488)	(0.0229)	(0.0506)
industry	-0.0164	0.0959	0.0319	-0.0300	0.0633	-0.0262	-0.00401	0.00327	0.0163	0.0639	-0.0143	-0.176*	-0.0266	-0.154
	(0.0388)	(0.111)	(0.0363)	(0.0913)	(0.0581)	(0.160)	(0.0410)	(0.122)	(0.0415)	(0.151)	(0.0228)	(0.0939)	(0.0170)	(0.110)
union	0.00103	-0.0517	0.00387	-0.0806**	0.0194**	-0.153***	0.00426	-0.0200	0.0139*	-0.00655	0.00301	-0.0434	-0.000501	0.0282
	(0.00802)	(0.0387)	(0.00433)	(0.0339)	(0.00977)	(0.0510)	(0.00690)	(0.0357)	(0.00811)	(0.0546)	(0.00309)	(0.0319)	(0.00313)	(0.0299)
public sector	-0.00332	-0.0833**	-0.0160**	-0.0450*	-0.00489	0.0157	-0.0138**	-0.0171	-0.00731	0.0157	-0.00945*	0.0132	-0.0168***	0.0234
	(0.00318)	(0.0354)	(0.00789)	(0.0254)	(0.00819)	(0.0347)	(0.00557)	(0.0249)	(0.00725)	(0.0398)	(0.00506)	(0.0239)	(0.00539)	(0.0218)
Constant		-0.354		-0.138		0.182		-0.335		-0.145		0.407**		0.402*
		(0.273)		(0.205)		(0.265)		(0.216)		(0.313)		(0.188)		(0.206)
Observations	6,294	6,294	15,383	15,383	10,737	10,737	13,314	13,314	15,388	15,388	15,825	15,825	15,003	15,003

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Omitted category Single, African, from western cape, non unionized, with primary school education or lower, in an elementary occupation in the manufacturing sector in the public sector

Table 14: Logit model results for constructing the re-weighting factor

VARIABLES	OHS1994	OHS1997	OHS1999	LFS01:2	LFS03:2	LFS06:2	QLFS2011:4	QLFS2014:4
Less20yrs	0.0275*	0.0102	-0.00381	0.00529	0.00384	-0.0236	-0.0186*	-0.0312***
	(0.0144)	(0.0131)	(0.0152)	(0.0119)	(0.0139)	(0.0154)	(0.0101)	(0.0115)
Less30yrs	0.0184	-0.0233*	-0.0450***	-0.0178	-0.0470***	-0.0509***	-0.0417***	-0.0358***
	(0.0151)	(0.0140)	(0.0168)	(0.0127)	(0.0148)	(0.0174)	(0.0108)	(0.0120)
more30yrs	0.0352**	0.00449	-0.0379**	-0.0174	-0.0310**	-0.0662***	-0.0687***	-0.0792***
	(0.0154)	(0.0148)	(0.0173)	(0.0132)	(0.0155)	(0.0184)	(0.0115)	(0.0127)
less_than_matric	0.0323***	-0.0256***	-0.0160	0.00737	0.00989	-0.0452***	-0.00500	-0.0256**
	(0.0117)	(0.00950)	(0.0114)	(0.00941)	(0.0106)	(0.0125)	(0.0110)	(0.0118)
Matric	0.0827***	0.0101	-0.00923	0.0528***	0.00516	-0.0117	-0.0178	-0.0294**
	(0.0181)	(0.0143)	(0.0171)	(0.0127)	(0.0144)	(0.0166)	(0.0128)	(0.0140)
Tertiary	0.130***	-0.00342	-0.00250	0.00915	-0.0331	-0.0383	-0.0376**	-0.0658***
	(0.0253)	(0.0190)	(0.0242)	(0.0188)	(0.0207)	(0.0239)	(0.0152)	(0.0176)
Managers	0.308***	0.125***	0.295***	0.266***	0.169***	0.152***	0.136***	0.118***
	(0.0312)	(0.0215)	(0.0281)	(0.0282)	(0.0293)	(0.0351)	(0.0173)	(0.0196)
professionals	0.0292	0.0296	0.158***	0.143***	0.151***	0.105***	0.0789***	0.0973***
	(0.0290)	(0.0190)	(0.0303)	(0.0251)	(0.0283)	(0.0345)	(0.0174)	(0.0238)
Technical/ass.profess	0.0180	0.0181	0.0587***	0.0932***	0.0582***	0.0289	0.0265**	0.0176
	(0.0217)	(0.0166)	(0.0209)	(0.0162)	(0.0183)	(0.0219)	(0.0128)	(0.0155)
clerks	-0.0781***	-0.136***	-0.0377*	-0.0707***	-0.0562***	-0.129***	-0.133***	-0.140***
	(0.0206)	(0.0182)	(0.0208)	(0.0164)	(0.0177)	(0.0222)	(0.0136)	(0.0150)
services	0.148***	0.166***	0.197***	0.171***	0.150***	0.112***	0.112***	0.0703***
	(0.0160)	(0.0134)	(0.0158)	(0.0125)	(0.0137)	(0.0157)	(0.0106)	(0.0112)
Skilled agriculture	0.308***	0.163***	0.420***	0.565***	0.130***	0.109*	0.0394	0.0814
	(0.0457)	(0.0220)	(0.0216)	(0.0276)	(0.0451)	(0.0639)	(0.0495)	(0.0502)
craft	0.368***	0.219***	0.308***	0.308***	0.343***	0.276***	0.333***	0.327***
	(0.0189)	(0.0139)	(0.0173)	(0.0141)	(0.0163)	(0.0177)	(0.0144)	(0.0148)
Machine operators	0.288***	0.328***	0.339***	0.288***	0.302***	0.267***	0.284***	0.275***
	(0.0159)	(0.0144)	(0.0160)	(0.0130)	(0.0149)	(0.0175)	(0.0142)	(0.0143)
Agriculture sector	0.401***	0.229***	0.224***	0.336***	0.303***	0.265***	0.246***	0.257***
	(0.0102)	(0.0115)	(0.0122)	(0.0110)	(0.0114)	(0.0143)	(0.0158)	(0.0153)
Mining sector	0.531***	0.528***	0.554***	0.600***	0.520***	0.522***	0.385***	0.297***
	(0.0393)	(0.0268)	(0.0316)	(0.0233)	(0.0308)	(0.0332)	(0.0258)	(0.0238)
Utilities	0.475***	0.394***	0.422***	0.349***	0.272***	0.290***	0.314***	0.186***
	(0.0562)	(0.0421)	(0.0713)	(0.0379)	(0.0423)	(0.0508)	(0.0473)	(0.0478)
Construction	0.389***	0.469***	0.436***	0.444***	0.371***	0.382***	0.360***	0.339***
	(0.0309)	(0.0229)	(0.0271)	(0.0211)	(0.0235)	(0.0301)	(0.0159)	(0.0158)
Trade	0.0805***	0.0916***	0.0777***	0.119***	0.0951***	0.106***	0.102***	0.106***
	(0.0143)	(0.0129)	(0.0152)	(0.0121)	(0.0135)	(0.0156)	(0.0102)	(0.0113)
Transport	0.453***	0.313***	0.331***	0.310***	0.305***	0.306***	0.287***	0.312***
	(0.0233)	(0.0222)	(0.0280)	(0.0198)	(0.0234)	(0.0269)	(0.0161)	(0.0181)
Finance	0.136***	0.173***	0.113***	0.135***	0.173***	0.175***	0.153***	0.176***
	(0.0223)	(0.0180)	(0.0192)	(0.0147)	(0.0160)	(0.0202)	(0.0104)	(0.0114)
Manufacturing	0.133***	0.155***	0.118***	0.150***	0.121***	0.170***	0.161***	0.145***
	(0.0155)	(0.0129)	(0.0175)	(0.0129)	(0.0159)	(0.0192)	(0.0119)	(0.0145)
married	0.0815***	0.122***	0.118***	0.115***	0.120***	0.105***	0.0871***	0.0975***
	(0.00912)	(0.00790)	(0.00938)	(0.00746)	(0.00828)	(0.0104)	(0.00692)	(0.00747)
coloured	-0.135***	-0.0710***	-0.0601***	-0.0544***	-0.0832***	-0.0657***	-0.0352***	-0.0178
	(0.0133)	(0.0140)	(0.0185)	(0.0143)	(0.0146)	(0.0193)	(0.0121)	(0.0124)
Indian	0.0301	0.0324	-0.0295	0.0222	-0.00373	-0.0370	0.0187	0.0444*
	(0.0204)	(0.0228)	(0.0288)	(0.0213)	(0.0245)	(0.0322)	(0.0197)	(0.0258)
White	-0.0929***	-0.0629***	-0.0990***	-0.0823***	-0.0954***	-0.0461**	-0.0338***	-0.0292**
	(0.0153)	(0.0155)	(0.0189)	(0.0135)	(0.0173)	(0.0228)	(0.0114)	(0.0134)
Observations	18,682	22,247	15,967	22,323	19,251	22,603	25,426	22,857

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1, Source: Author's own from PALMSV3.1

Notes: Omitted groups: African, single with primary education, from western cape, elementary worker in the manufacturing sector.

4 A COHORT ANALYSIS OF THE GENDER WAGE GAP

4.1 INTRODUCTION

Results from the cross-sectional analysis of the gender wage gap in chapter three showed that the mean gender wage gap in South Africa narrowed significantly from 0.34 log points (about 40%) in 1993 to 0.15 log points (about 16%) in 2014. This decline was attributed to a declining unexplained gender wage gap. Several factors may have contributed to the declining unexplained wage gap including improved education for women, changes in behavioural and cultural norms, and labour market legislation. Women now make up a larger percentage of individuals with tertiary education at about 55 percent in 2015, up from about 40 percent in 1993. There has been an increase of women in the labour market and an improvement in the share of women in professional occupations (see figure 1). For example, the share of women managers and legislators almost doubled in the period 1993 to 2015.

Figure 22 below shows that the proportion of females in employment[30] has been increasing over time with some fluctuations. The figure shows that for both men and women, there was a decline in the proportion of those in employment between 2000 and 2003. This trend however reversed and there was an increase in employment between 2003 and 2008. There was then a drastic drop in employment between 2008 and 2010 for both men and women. This drop is mostly due to the global financial crisis of 2007-2008, the effects from which South Africa was not exempt. What is interesting, is that the drop in employment was more severe for men than for women for all age groups. The most affected group, however, is the group of males aged between 25 and 34. The proportion employed for this group dropped from about 0.68 in 2008 to about 0.58 in 2015. For the same group of women, the proportion of those employed declined from 0.46 in 2008 to about 0.4 in 2010 but then increased to about 0.44 by 2015. Therefore, even though the proportion of men in employment is higher than that of women for all age groups, relative to men, women made more relative gains in employment in the period between 2010 and 2015.

Given the developments described above, it is plausible to assume that more recent generations of women are faring better in the labour market than their predecessors. A short coming of cross-sectional

[30] This refers to both wage employment and self-employment.

analysis however, is that cross-sectional data by its nature compares individuals from different age groups and generations at one point in time. Earnings have a distinct life cycle profile in that they are low at younger ages, increase as an individual approaches middle age and decline later in life as one approaches retirement (the so called 'hump' shape) (Deaton 1997; Mincer 1974). Additionally, earnings trajectories of successive generations differ because earnings respond to changes in the general macro-economic environment (Deaton 1997).

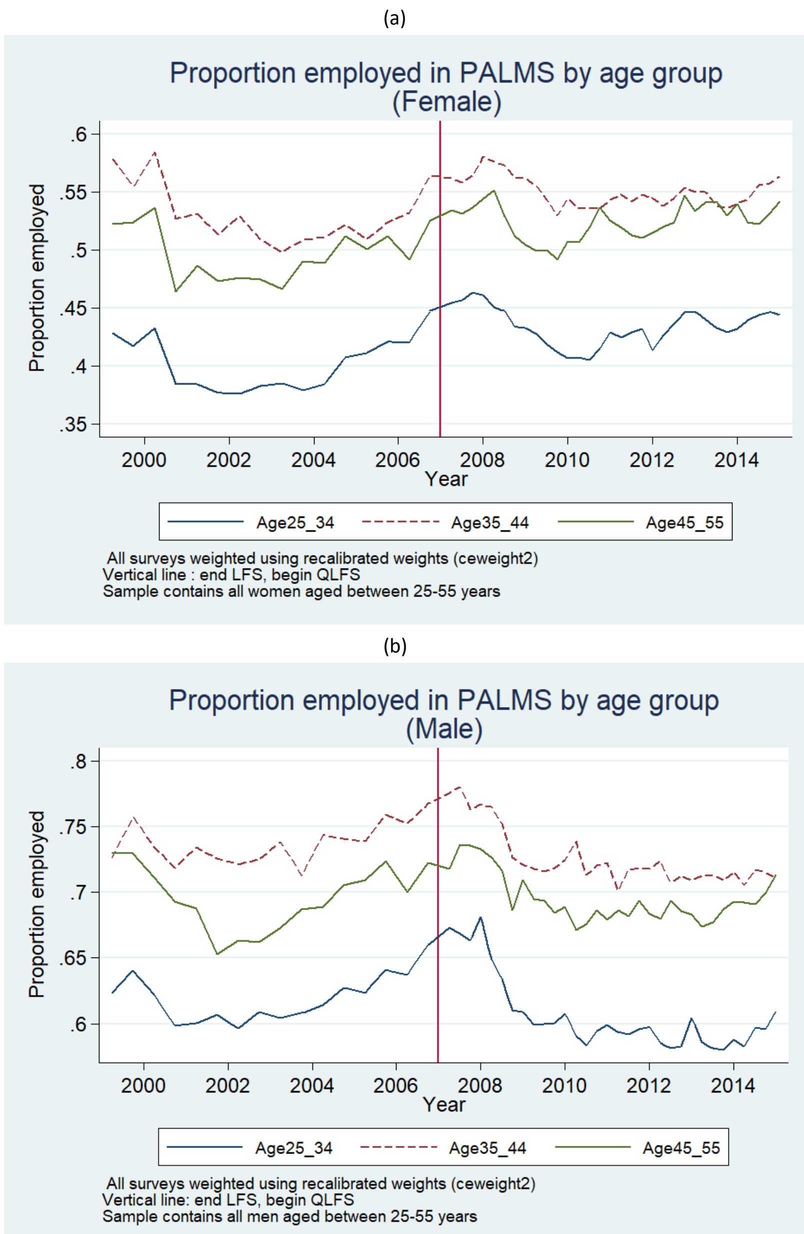


Figure 22: Proportion employed in PALMS by age group

To answer the question of whether the current generation is faring better in the labour market in terms of earnings and the gender wage gap, long spanning panel data that follows the same individuals over their entire life cycle is required.

Long-spanning panel data are however scarce in most countries. South Africa does have a nationally representative panel; the National Income Dynamics Study (NIDS) which tracks individuals of all ages in South Africa and collects data on income dynamics of South Africans. The NIDS began in 2008 with a nationally representative sample of over 28,000 individuals in 7,300 households across the country. These individuals are reinterviewed every two to three years. At the time of writing this thesis however, the NIDS dataset only had 4 waves of data spanning only 7 years. In this chapter, we therefore utilise synthetic cohort data constructed from repeated cross sections to analyse dynamic changes in the South African labour market over time following Deaton (1985). We define a cohort as a group of individuals who share the year of birth. Organizing data into a synthetic cohort is useful even in the presence of real panel data as it allows for the analysis of longer periods (Gardes et al. 2005).

This chapter makes the following contributions, we construct cohort data which allows us to look at long term trends of male and female earnings. This is important, especially for the African sub-population, to assess the effects of labour market transformation in post-apartheid South Africa. The topic of labour market transformation and its effects is an important one given the history of apartheid where black South Africans were denied proper education which limited them from competing for skilled occupations. Some jobs in the labour market were also reserved for the white population. Labour market transformation since 1994 has been focused on eliminating these inequalities. There is a need therefore to investigate whether younger cohorts who joined the labour market after the end of apartheid have experienced any improvements in the labour market.

Second, we use the cohort data to analyse the gender wage gap. We ask the questions; How do male and female wage trajectories differ over the life-cycle? Does the wage gap differ across cohorts indicating a generational effect? Has the within cohort gender wage gap declined over time? We expect that women who joined the labour market after the Employment Equity Act was enacted, have experienced a lower wage gap than their predecessors. The Employment Equity Act was enacted in 1998 and it was the first legislation that was specifically aimed at promoting equal opportunity and fair treatment in employment through the elimination of unfair discrimination.

Our analysis of life cycle trends show that younger cohorts of men and women have experienced a rise in wages over time which we attribute to improved human capital characteristics. More recent cohorts of women have access to better education and higher paying jobs compared to women who were in the labour market during apartheid. This has led to an increase in the average wages of women leading to the closure of the gender wage gap for the youngest cohorts.

The rest of this chapter is organized as follows, in section 4.2 we review empirical literature while the data and methods used in this analysis are presented in section 4.3. In sections 4.4 and 4.5 we discuss errors in measurement and selection bias respectively. Section 4.6 presents the results. Finally, we discuss the results and present our conclusion in section 4.7.

4.2 EMPIRICAL LITERATURE

Both locally and internationally, studies report a declining gender wage gap over time (Polachek 2006; Shepherd 2008; Blau & Kahn 2000). However, the decline of the gender wage gap over time could be due to recent cohorts experiencing a better labour market environment (period effects). It could also be the case that more recent cohorts have better control over their reproductive health (can delay marriage and the birth of their first child) (cohort effects and age effects). It is challenging to distinguish between these three effects. Labour market legislation, such as anti-discrimination policies in the labour market, can lead to a decline in the gender wage gap over time (period effects). However, it can also lead to cohort effects in that the cohorts that join the labour market after the legislation will have a completely different earnings trajectory than their predecessors.

For post-apartheid South Africa especially, long term trends on earnings and the gender wage gap are important because there is need to take stock of the progress in labour market transformation. Equal employment legislation has contributed to changing female attitudes towards the labour market leading to increased participation (Ntuli & Wittenberg 2013). Longer attachment to the labour market has a narrowing effect on the gender wage gap. Anti-discrimination legislation might affect different cohorts differently. Cohorts that finished school after the enactment of the Employment Equity Act may have experienced better employment opportunities and better starting wages. Similarly, younger cohorts may have experienced better education than their predecessors. In post-apartheid South Africa, younger cohorts, especially from the African race group, have acquired more education than their predecessors. Additionally, there are more women with tertiary education among the recent cohorts. Minimum wage legislation has been found to have a positive effect on wages especially for women in low paying occupations (Hertz 2005).

Macro-economic conditions, such as the shedding of jobs in the manufacturing sector (Bhorat & Rooney 2017) which would mostly affect low skilled men, can lead to period effects on the gender wage gap. The effect comes from not necessarily the improvement of women’s conditions but from the deteriorating labour market outcomes for men. In the United States, the decline in manufacturing jobs which were disproportionately male jobs contributed to the narrowing of the gender wage gap in the 1980s (Blau & Kahn 2016). Similarly, decline in trade unionisation can also lead to period effects on the gender wage gap. Unionised workers earn higher wages and men are more likely to be unionised. However, in the recent past we have seen a decline in the rates of unionised men over time and they seem to be converging to those of women. This implies a lower wage gap for the recent cohorts of women. Declining marriage rates for recent cohorts is another factor that could lead to period effects on the gender wage gap. These arguments follow from the human capital explanations of the gender wage differentials in the labour market (Mincer & Polacheck 1974).

Like earnings, the gender wage gap displays life cycle effects (Tyrowicz et al. 2015). The wage gap is smaller at the onset of careers, it then increases as women drop out of the labour market due to child birth and care and then may decline later as women return to the labour market when children are older (Goldin 2014; Polachek 1975a). Bertrand et al. (2010), looking at careers of MBA graduates in the United States, find that the graduates display similar earnings at the start of their careers, but that the male and female earnings trajectories diverge after several years in the labour market. They attribute this divergence to lower weekly hours for women and work discontinuity which they attribute to motherhood.

In a cross-sectional analysis however, it is difficult to assess the correct effect of age on the gender wage gap as cross-sectional data only gives information at a point in time. Drolet (2011) investigating the gender wage gap in Canada, finds that within-cohort the gender wage gap widens very little over time as compared to when looking at the gender wage gap across different age categories in cross-sectional data. This is because in a cross-section the comparison is of individuals in different stages of their lives.

She concludes that cross-sectional analysis tends to overstate the correlation between the wage gap and age (Drolet 2011).

The closure of the gender gap in education also implies that recent cohorts expect to stay in the labour market longer and therefore are investing more in education which should increase their wages and reduce the wage gap (Polachek 1975b). Similarly, medical and technological advancements relating to fertility also define a different labour market experience and expectations for recent cohorts (Goldin & Katz 2002). The narrowing of the gender wage gap in recent years has been attributed to discoveries such as the pill. Goldin & Katz (2002) find that women born after the introduction of the birth control pill have a more constant participation in the labour market and display a different earnings trajectory than those born before.

Life cycle analysis requires long running panel data where individuals are followed over their life cycle. In the developing countries context, where absence of such long-spanning panel data is common, Deaton (1985) suggested the use of pseudo panels or cohort data from repeated cross sections. Cohorts are defined by year of birth and allow us to follow groups of people from one survey year to the next. This makes cohort data similar to panel data except that groups of people and their average characteristics are followed and analysed instead of individuals.

Deaton (1997) outlines several advantages of cohort data created this way. First, is that since cohort data are constructed from new surveys every year, cohort data does not suffer from attrition like true panel data. Secondly, since averaging usually reduces the effects of measurement error, an analysis using cohort data is less susceptible to measurement error since the unit of analysis is an average (Deaton 1997, p.120). Third, cohort data allows one to analyse the relationship between household behaviour and national aggregates.

The pseudo-panel data approach has been applied in literature to study a range of topics including life cycle savings (Attanasio 1993; Browning & Lusardi 1996; Deaton & Paxson 1994) consumption (Deaton 1997; Meng et al. 2014), marriage trends (Kumchulesi 2010), trends in education, employment and earnings (Branson et al. 2013; Burger & Von Fintel 2009; Branson & Wittenberg 2007; Warunsiri & McNown 2010) and labour force participation (Contreras et al. 2005; Devereux 2007; Beaudry & Lemieux 1999).

In the South African literature, Burger & Von Fintel (2009) decompose unemployment, labour force participation and wages in South Africa into period, cohort and age effects. Their results indicate that the higher unemployment rates faced by the youth are mainly due to the fact that they entered the labour market more recently, and not due to their age. They also attribute increased labour supply among the more recent entrants to higher educational attainment and changes in household formation decisions over time. Similarly, Branson et al. (2013) use cohort analysis to study the changes in the distribution of education across birth cohorts and how this relates to the probability of employment and the distribution of earnings among white and black men. Their results showed that younger cohorts increasingly faced worse labour market conditions than their predecessors. This result motivates our study in that it is important to investigate whether recent cohorts of women have experienced better labour market conditions than their predecessors and what the implication is for the gender wage gap.

Pseudo panels have also been applied in the study of earnings differentials. For example, Fitzenberger & Wunderlich (2002) examined period and cohort effects on the gender wage gap in Germany and found that there were no cohort effects for females. Additionally, they found that females exhibited lower wage growth through the lifecycle. Smyk et al. (2014) investigate age productivity wage differentials in talent occupations in Poland and find diverging age-earnings profiles between men and women in talent occupations. Campbell & Pearlman (2013) using data from the current population survey in the United States attribute the narrowing of the gender gap to cohort effects. For cohorts born after 1950, they attribute the decline of the gender wage gap to the decline of male wages. The fact that life cycle trends in earnings differ by gender and by country and that the contribution of age, period and cohort effects to the gender wage gap also differs by country further motivates this analysis.

In South Africa, Grün (2003) studied racial and gender wage differentials using pseudo panel data consisting of OHS 1995, OHS 1997 and OHS 1999 data and found that the gender wage gap in South Africa exhibited cohort effects. Interestingly, for the African sub-sample, she found that wages seemed to diverge for the older workers and that, although not significant, there was a negative gender wage gap in 1995. The author found this result to be unusual but attributed the result to younger cohorts. For white workers the author found that the greatest wage differential was among middle-aged workers and attributed this trend to the possibility of women dropping out of the labour market for family reasons. Decomposing age, period and cohort effects of earnings for African women, Grün finds that older cohorts benefit from generational effects suggesting that those born earlier earn higher wages than younger cohorts. This result is contradictory to findings in literature which show that younger cohorts who may benefit from more (better) education earn higher wages than older cohorts (Deaton & Paxson 1994). Grün (2003, p.16) did however term her results controversial and considering that the study only utilised 3 cross-sectional surveys to construct cohort data, she concluded that more cross-sections are necessary for more conclusive results.

This study contributes to the gender wage gap literature by explicitly examining cohort effects and period effects on the gender wage gap in a developing country context. We also, exploit this methodology to model the life cycle trends in earnings for men and women in South Africa.

4.3 DATA AND METHODS

4.3.1 The Data and Measures

In this chapter, we exploit nationally representative cross-sectional data to construct a pseudo panel data set. We continue to use the PALMS dataset. Although the PALMS dataset contains cross-sectional surveys collected since 1993, for the current analysis we only utilise data from the September rounds of the Labour Force Surveys (LFS) (2000-2007) and quarter three of the Quarterly Labour Force Surveys (QLFS) from 2008 to 2014. We exclude the October household surveys from this analysis because of the data quality issues afflicting the OHSs including the jump in labour force participation that was concentrated between 1998 and 2000 (see figure 2). As discussed in section 1.3, part of this jump has been attributed to better data collection and this might affect the within cohort trends. Additionally, restricting the sample to the September rounds of the LFSs and quarter three of the QLFSs controls for seasonality.

We restrict our sample to African respondents aged between 25 and 50 to ensure that we have enough observations and to reduce heterogeneity within cohort cells. The age restriction ensures that we eliminate individuals who are still in school and those who have retired from the labour market. In this study we are interested in analysing the effect of age, period and birth cohort effects on wages for men and women separately and on the gender wage gap. For the male and female wage trajectories, the dependent variable is the log of real hourly earnings[31] whereas for the gender wage gap the dependent variable is the gender wage gap calculated as $logmalewage - logfemalewage$ for each birth cohort. To deal with the issue of earnings given in brackets and outliers we utilize bracket weights and the outlier detector variable provided in PALMS.

Explanatory variables include age, period and cohort variables which are modelled as dummy variables so as not to impose any particular relationship between these variables and the wage gap (earnings). In all the regressions we control for education, marital status and geographical location. As additional variables, we therefore include a four-category education variable, a marital status dummy[32] and a 9-category province variable.

4.3.2 Construction of Cohort Data

We define a cohort as individuals who share the same birth year where birth year is calculated as $surveyyear - age$. Since we have restricted age to between 25 and 50, the youngest cohort was born in 1989 and is aged 25 in year 2014 while the oldest cohort was born in 1950 and is aged 50 in year 2000. As the main interest in this analysis is the gender wage gap, to create the pseudo panel data, we average variables over individuals who share the same birth year and gender. With 26 age groups, 2 groups for gender and 13 period groups (15 years from 2000-2014 but less 2008 and 2009 as there was no earnings information for these two years) we end up with 676 age-cohort cells (See table 17 and 18 in the appendix).

Average earnings of each cohort are tracked from the time the cohort is first observed in the dataset until they are last observed. For example, the cohort born in 1975 is first observed in the dataset in 2000 when they are aged 25 and tracked until they are last observed in 2014 when they are aged 39. For this cohort, average earnings at age 25 are calculated from the LFS 2000. This gives us the first data point of their wage distribution. The second data point is at age 26 and average earnings for this cohort are calculated from the LFS 2001. The next data point is at age 27 and average earnings are calculated from the LFS 2002. This process forms a line of connecting data points showing us the trend of the 1975 birth

[31] We use a real earnings variable provided in PALMS to calculate real hourly earnings by dividing real monthly earnings by monthly hours where monthly hours equal hours worked in the last week multiplied by average weeks in a month. We took average weeks in a month to be 4.333.

[32] where married=1 for married individuals and married=0 for everyone else

cohort's wage distribution until they are last observed at age 39 in 2014. Over time the wage path of different cohorts will become apparent and we expect that different cohorts will exhibit different patterns.

For all the descriptive graphs we utilize locally weighted scatter plot smoothing (LOWESS) (Cleveland 1979) to easily show relationships between variables over time. To ensure comparability and continuity between surveys we utilize cross entropy (CE) weights (Branson & Wittenberg 2014) also provided in PALMS. Cohort data allows us to see changes that have taken place over time by comparing different generations. We are able to track earnings of individuals born in the same year and hence clearly see the growth in earnings over their life cycle. For example, the earnings of individuals who are 25 years old in 2000 should be compared with those that are 30 years old in 2005 and not ones that are 30 years old in 2000 if we are interested in life cycle changes in earnings within a cohort. Individuals who are 30 years old in 2000 were born in 1970 while those that are 25 years old in 2000 were born in 1975. These two groups are from different generations and may have joined the labour market under different macro-economic conditions.

4.3.3 Age, Period and Cohort Effects Decomposition: Male and Female wages and the gender wage gap

Organizing repeated cross-sectional data into cohort data allows us to decompose age, cohort, and period effects on male and female wages and the gender wage gap. The decompositions are carried out using ordinary least squares method to determine the contribution of each effect. For changes in male and female wages over time we utilise equation (21) and (22).

$$lnW_{ct}^{f} = \beta_1^{f} + \beta_2^{f} X_{ct}^{f} + \theta_c^{f} coh_c^{f} + \gamma_y^{f} year_t^{f} + \delta_{ct}^{f} age_{ct}^{f} + \varepsilon_{ct}^{f} \tag{21}$$

$$lnW_{ct}^{m} = \beta_1^{m} + \beta_2^{m} X_{ct}^{m} + \theta_c^{m} coh_c^{m} + \gamma_y^{m} year_t^{m} + \delta_{ct}^{m} age_{ct}^{m} + \varepsilon_{ct}^{m} \tag{22}$$

for $c = 1, \ldots, C, t = 1, \ldots, T$

Where lnW_{ct} is the log hourly wage for group g where $(g = male, female)$, in year $t(t = 2000, 2001, \ldots, 2014)$ and for cohort c. coh_c^g is the birth cohort of group g, $year_t^g$ is the survey year in which cohort c in group g is observed and age_{ct}^g is the age in years of cohort c in year $t(t = 2000, 2001, \ldots, 2014)$. δ_{ct}, θ_c and γ_y represent the age, cohort and time effects and ε_{ct} is an independent zero mean random error term. X_{ct} represents the different average cohort characteristics that determine wages and the βs are coefficients of these characteristics.

We run two separate regressions as we are interested in answering the question of whether younger cohorts are experiencing better wage outcomes and whether there are differences in the age-earnings, cohort-earnings and period-earnings profiles between men and women. This is because returns to different characteristics may differ significantly between men and women. These differences will also help us understand the results from the gender wage gap regression as per equation (23).

To analyse changes in the wage gap over time we utilise equation (23).

$$lnGap_{ct} = \beta_1 + \beta_2 X_{ct} + \theta_c coh_c + \gamma_y year_t + + \delta_{ct} age_{ct}\, \varepsilon_{ct} \tag{23}$$

for $c = 1,\ldots,C, t = 1,\ldots,T$

Where $lnGap_{ct}$ is the gender wage gap[33] in year $t(t = 2000,2001,\ldots,2014)$ for cohort c and age_{ct}, δ_{ct}, θ_c, γ_t, X_{ct} and ε_{ct} are as described above.

4.3.4 Identification of Age, Period and Cohort Effects

The difficulty in estimating equations (21), (22) and (23), is the identification resulting from the perfect collinearity in these models. This is because $age + cohort = year$ and hence the design matrix, X containing age, period and cohort variables, is singular and therefore $(X^TX)^{-1}$ does not exist, making a unique solution to each of the effects impossible. In the literature several solutions to this problem have been proposed (Heckman & Robb 1985; Deaton 1997; Yang & Land 2006; Yang et al. 2008).

Heckman & Robb (1985) argue that models such as equation (23) can be justified because all the variables in the model are proxies for underlying unobserved variables that are not themselves linearly dependent. Age for example is a proxy for physiological variables and screening measures by employers and year effects are proxy variables for macroeconomic conditions. They argue that if one were to use one or more proxy variables as surrogates for the age, period or cohort variables then the identification problem would be solved (Heckman & Robb 1985).

An alternative method that has been suggested and used in the literature is the coefficients constraints approach (Mason & Fienberg 2012). This method requires one to place one or two additional equality constraints on the parameter vector so that the model is just or over identified. For example, this can be

[33] The wage gap is calculated as $logmalewage - logfemalewage$ for each birth cohort

done by assuming some categories of age groups, cohorts or time periods have identical effects on the dependent variable. For example, one can assume the coefficients of the first two periods to be equal therefore making the model just identified. The difficulty with this method comes in identifying the constraints which requires external information that might not be available (Yang & Land 2006). Other researchers have also opted for a linear restriction on period effects assuming they are equal and are included in the constant term (Russell & Fraas 2005).

Deaton & Paxson (1994) suggested a normalization procedure to identify the three effects that assumes that period effects sum to zero and are orthogonal to a time trend, forcing any time trend to appear as a combination of age and cohort effects and allowing the model to be just identified. Formally they impose the constraints: $\Sigma_t \pi_t = 0$ and $\Sigma_t t * \pi_t = 0$. Deaton (1997) however cautions that when only few cross-sectional surveys are available, this procedure might be dangerous as it will be difficult to separate trends from transitory shocks. He recommends at least 15 years of surveys.

A more recent approach is the Intrinsic Estimator proposed by Yang et al. (2008). The Intrinsic Estimator can be viewed as a regression of the y vector on the non-zero eigen-vectors of the $X^T X$ matrix or a principal component regression estimator that is transformed back to the coordinates of the original age-period-cohort space to obtain interpretable coefficients (Yang et al. 2008). This method has been applied in the literature for example by Branson et al. (2013). A critique of the Intrinsic Estimator by Luo (2013) is that similar to the coefficients constraints approach, it imposes a constraint on parameter estimation that is difficult to verify using empirical evidence or theory.

Since in the literature there is no agreed upon solution to the identification problem and that all the suggested methods have serious criticisms, we chose the Deaton-Paxson decomposition as it is the most commonly applied method in the literature (Grün 2003; Burger & Von Fintel 2009; Smyk et al. 2014). The disadvantage with this method is that averaging time effects to zero means that if there are any permanent effects over time, these effects will be added to cohort effects. This is to say that if there are wage differentials between earlier and more recent surveys, (for example caused by labour market legislation) this effect will be included in the cohort effect. Given the anti-discrimination labour market legislation in South Africa since 1994, we expect that the cohort effect from this model will be inflated.

To estimate equations (21), (22) and (23), we apply the Deaton & Paxson (1994) normalization procedure. To identify the three effects, we assume that wage growth and changes in the gender wage gap are attributable to age and cohort effects and that year effects will “capture cyclical fluctuations” that will sum to zero in the long run (Deaton 1997, p.126). We regress the gender wage gap on covariates, dummies for each age, dummies for each cohort and a set of $T - 2$ year dummies defined as

$$d_t^* = d_t - [(t-1)d_2 + (t-2)d_1] \tag{24}$$

4.4 ERRORS IN MEASUREMENT

According to Deaton (1997) cohort averages in a sample are different from true population averages as a result of sampling error and an errors-in-variables model is required for credible parameter estimates. Two suggestions for solving this problem according to Deaton (1997) are either an instrumental variables method where changes from previous years are used as instruments or by subtracting variances and covariances of the sampling errors from the variances and covariances of the sample cohort averages. Verbeek & Nijman (1993) however showed that if cohort cells contained about 100 people, sampling measurement error could be ignored. Additionally, if the individuals grouped in each age-cohort-period cell are homogeneous the number of individuals in each cohort can be smaller (Verbeek & Nijman 1993). In this analysis we deal with measurement error by ensuring that each cell contains individuals that are as homogeneous as possible by restricting our sample to Africans only. Additionally, out of 676 cells only 11 cells have less than 100 individuals (see table 17 and 18 in the appendix) and therefore we assume that measurement error is not a major problem in this analysis.

4.5 SELECTION ISSUES

An important issue that may affect our findings is the issue of selection. It is well documented that some women, especially the highly educated women may take time off from the labour market and return later in life (Contreras et al. 2005; Polachek 1975a). This may result in a dip in female earnings in their midlife and an increase after childbearing is complete. We would therefore expect the gender wage gap to be narrow at the start of woman’s career, to expand towards the middle and start narrowing towards the end of their career. New research also shows that the age at which women have their first child matters for the gender wage gap. Chung et al. (2017) show that women who give birth to their first child before the age of 25 and after the age of 35 suffer a smaller shock to their life time earnings and are more likely to close the parental penalty.

A key assumption of a cohort analysis is that cohorts are closed. That is, individuals cannot move from one cohort to another. However, labour market cohorts are dependent on someone being in the labour market and we know that people move in and out of the labour market. The analysis is therefore vulnerable to selection as women of childbearing age could be dropping out of the labour market to have and raise children. If for example, the kind of women who leave the labour market are all low earning women and of a certain age, then what might look like an increase in wages for a certain cohort may not be a true increase in wages but rather a function of selection in the remaining sample. Women with low education (less than high school) are more likely to have their first child earlier in life (in their 20s) whereas women with higher education are more likely to have their first child later in life (in their late thirties) (Chung et al. 2017).

Below we look at the trends in employment and labour force participation to assess potential selection in our sample. What is important to consider in this analysis is whether selection works differently for men and women and for different age groups, period groups or cohorts.

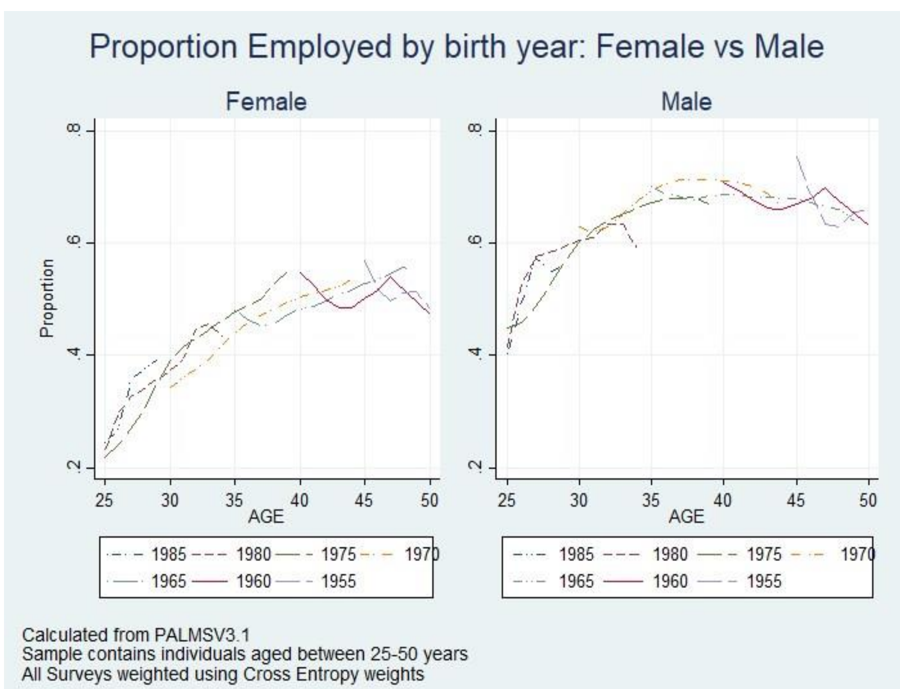


Figure 23: Proportion of the labour force in employment over time

Figure 23 shows the proportion of the labour force that is employed by gender over the life cycle using the PALMS dataset. Each line segment represents a specific birth cohort and each point on a particular line is the proportion of either men or women employed at a specific age.

As an example, the cohort born in 1980 is first observed in 2005 when it is aged 25 and is tracked until 2014 when it is aged 34. The employment rate for men in this cohort when it is first observed is about 40%, whereas that of women is about 22%. For men the proportion employed grows to almost 60% by 2014 and that of women rises to about 42%.

The figure shows a persistent employment gap between men and women. What is also apparent from the figure is that the proportion employed increases with age for both men and women. The employment trends for men and women are similar, however, the proportion of employed men seems to level off around age 35 possibly signalling early retirement. The fact that these trends are similar between men and women is an indication that even if there could be selection happening between different age groups, it looks like the selection is not changing over time. So, our analysis will pick up the right cohort and period effects.

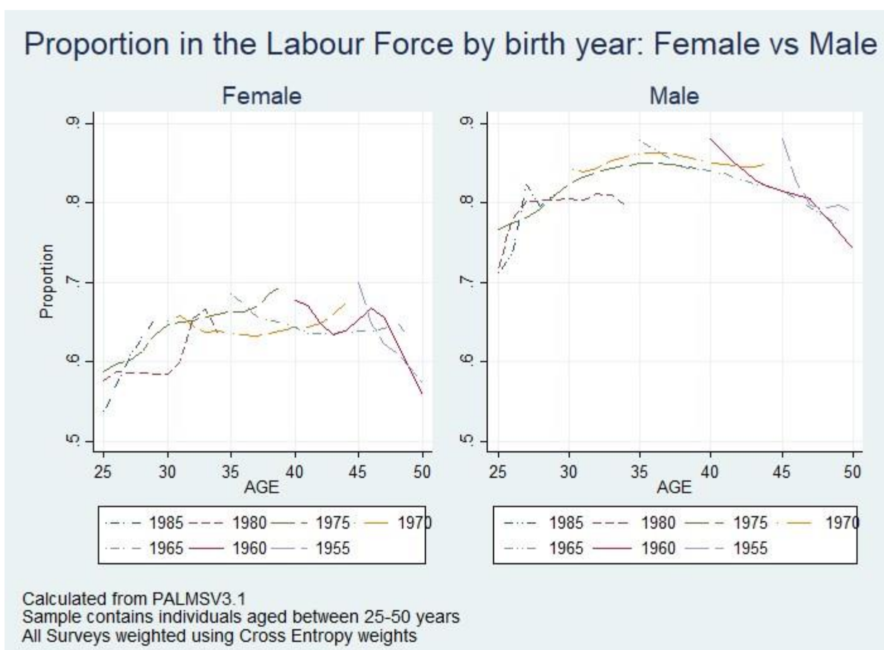


Figure 24: Labour Force Participation by Cohort over time

Figure 24 plots the labour force participation rate by cohort and age. Like the employment figure above, the participation rate increases with age up to a point. For both men and women, participation proportions for different cohorts overlap showing that different cohorts observed at the same age have similar participation rates. The above figure shows that peculiarly, labour force participation starts to decline after age 35 for older cohorts. This suggests that there is another reason other than motherhood and caring for family that is leading to the decline in participation rates after this age. Looking at cross sectional surveys over the analysis period, the same trend is evident.

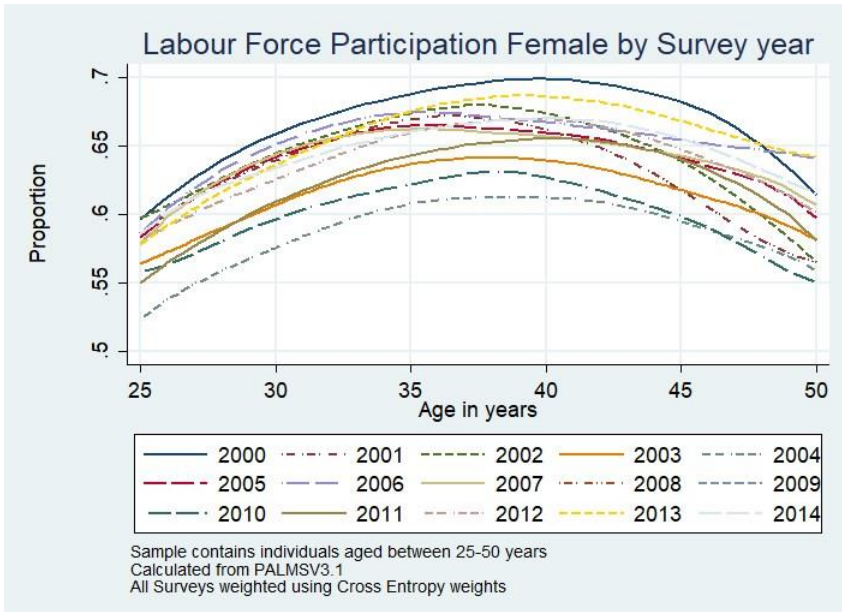


Figure 25: Female Labour Force Participation by period and age

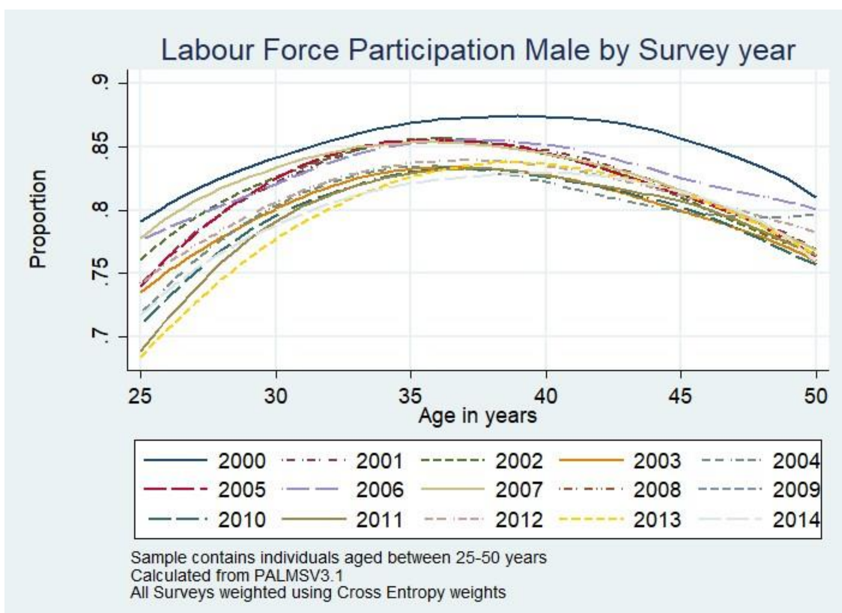


Figure 26: Male Labour Force Participation by period and age

Looking at figures 25 and 26, each line segment represents a different survey year and plots the participation rate by age. We see that year 2000 was an outlier in that it displays much higher labour force participation rates for both men and women than other survey years. This trend is well documented in the literature (Casale & Posel 2002). For women, survey 2004 seems to be another outlier but in the opposite direction, displaying visibly lower participation rates. What is important for this discussion, however, is that the figures show that for both men and women, labour force participation rates start declining around age 40 which is still in the prime working age years. This supports the discussion above that there seems to be an early decline in labour force participation before the retirement age.

4.6 RESULTS

4.6.1 Descriptive Statistics

To help us understand the resulting trends from this analysis, we look at some cohort characteristics that are important for wages and the gender wage gap. Looking at the trends in average years of education by cohort and gender (see figure 27), it is apparent that younger cohorts of men and women are more educated than older cohorts. Also evident is that younger cohorts of women have slightly more years of education, on average, than men.

Figure 28 shows that there is a decline in the proportion of individuals that report being married as we move from older to younger cohorts, pointing to a generational change. Comparing the cohorts of men born in 1965, 1970 and 1975, at age 35, the proportion of employed men who are married was about 0.7 in 1965, whereas for the cohort born just five years later in 1970 this proportion had dropped to about 0.62, and for the cohort born in 1975 this proportion had dropped further to about 0.58.

Figure 29 shows that older cohorts were more likely to be in a union than younger cohorts. Also, clearly visible is that men were more likely to be in a union than women, but this gap has been closing over time.

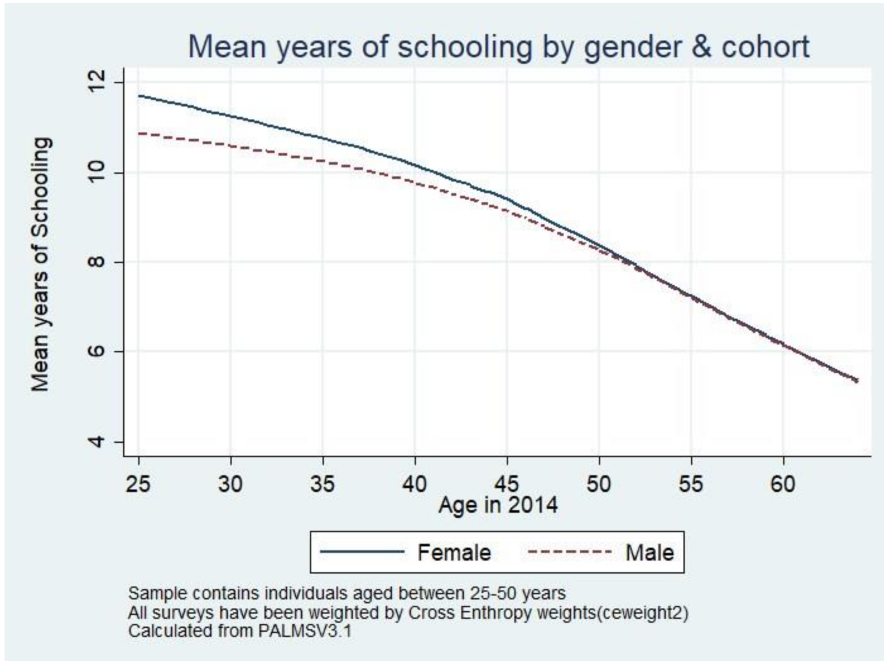


Figure 27: Mean years of schooling by gender

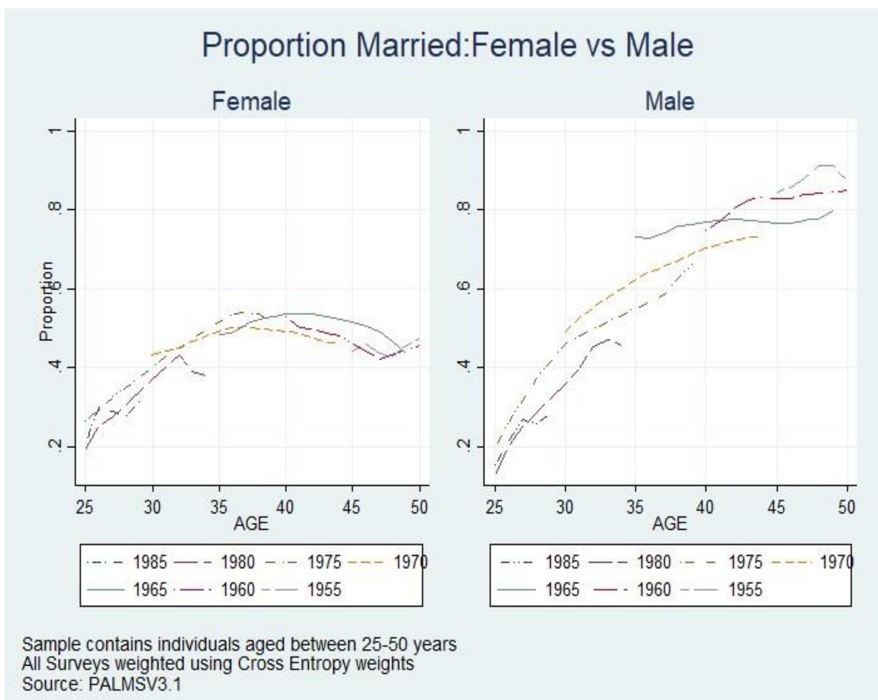


Figure 28: Proportion Married by birth cohort and gender

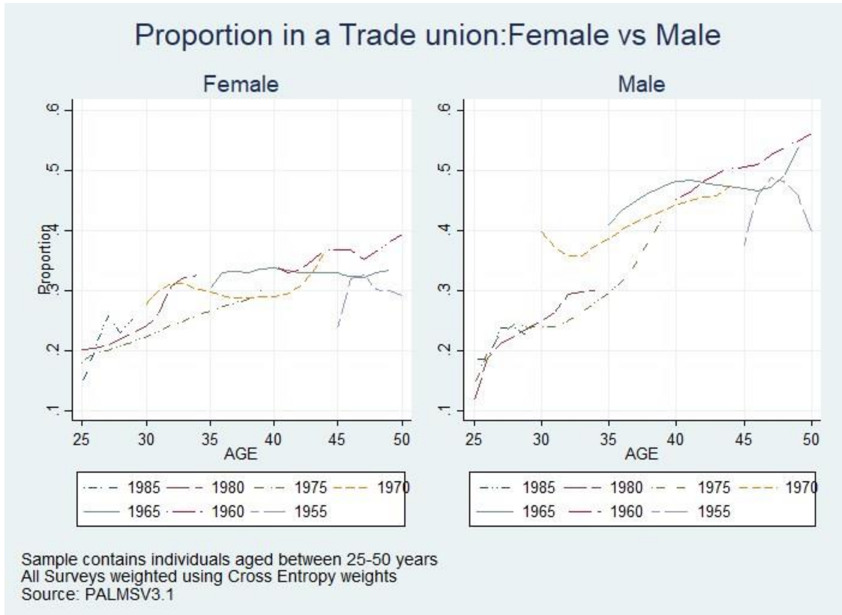


Figure 29: Proportion in a trade union by birth cohort and gender

4.6.2 Life Cycle Patterns of Earnings and the Gender Wage Gap

Figure 30 plots the average real log hourly earnings for selected birth cohorts separately for men and women. The way to read this figure is that every connected line represents a particular cohort's average earnings from the time that cohort is first observed in the dataset to when it is last observed. For example, the cohort born in 1975 is first observed in the dataset in 2000 when it is aged 25 and continues to be observed until it is aged 39 in 2014. The trends show that earnings increase with age for both men and women and that they increase faster for younger cohorts. Earnings trajectories for younger cohorts are above the earnings trajectories of older cohorts, suggesting that younger cohorts have benefited from higher wages. Take the female cohorts born in 1980 and 1975. At age 30, the cohort born in 1975 was earning about 1.6 log points whereas the cohort born five years later in 1980 was earning about 1.9 log points at age 30.

Comparing the male and female figures, for every birth cohort, male earnings are higher than female earnings from the time the cohorts are first observed until they are last observed. This gap has however been declining over time. Take the 1955 male cohort, their earnings when first observed at age 45 are almost 2 log points compared to the 1955 female cohort whose earnings are about 1.5 log points, a wage gap of about 0.5 log points. Comparing this with the 1980 female and male cohorts, the wage gap seems to be declining. The gender wage gap for the 1980 cohort at age 30 is about 0.2 log points with men earning about 2.1 log points and women earning about 1.9 log points. Looking at the between cohort gap for women and men, it seems that women have had more relative gains in wages than men, leading to this decline in the wage gap.

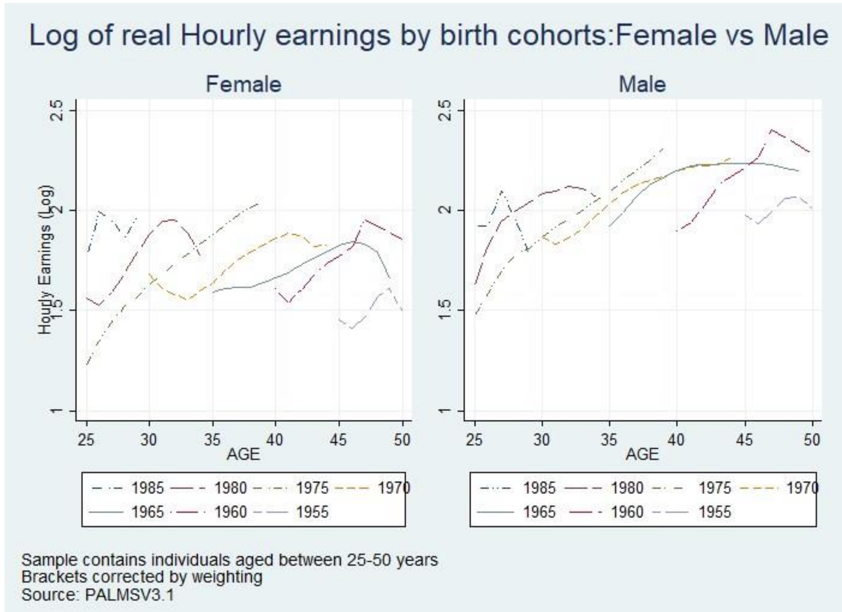


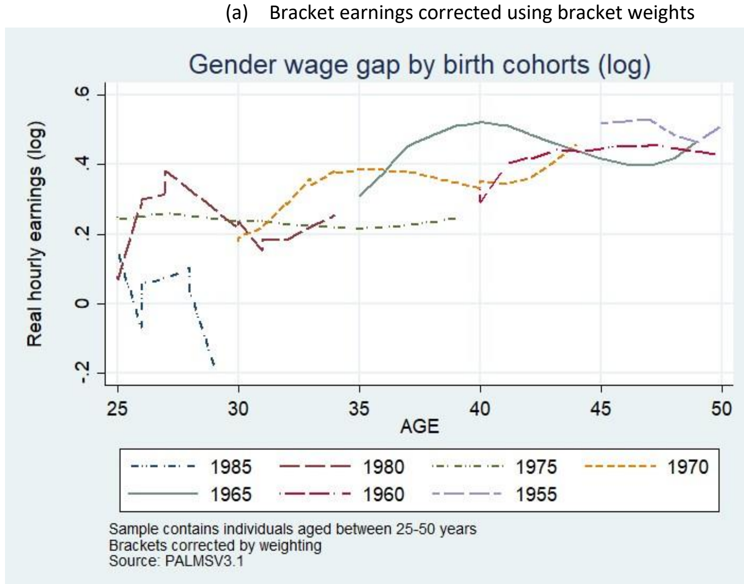
Figure 30: Life cycle real earnings for selected cohorts by gender

It is illustrative to take a closer examination of the 2 panels in figure 30. A 25-year-old female in 2010 was earning about 0.5 log points higher than a 50-year-old female in 2005. This could be because a 25-year-old female in 2010 is characteristically different from a 50-year-old in 2005. More recent cohorts of women have more education and are more likely to be working in a skilled or semi-skilled occupation. The 50-year-old most likely has less education and works in an elementary occupation. The male panel however shows that a 25-year-old male in 2010 was earning almost as much as a 50 year old in 2005. This is interesting and shows that women have had more relative gains in wages than men.

Looking at figure 30 we can see the limitations of analysing earnings using cross sectional data. From the figure we can see a combination of cohort, period and age effects, which are difficult to separate. Cohort effects can be seen by the fact that each cohort trajectory is above the previous cohorts. Age effects can be seen by the fact that earnings increase as individuals age. The fluctuations in each cohort trajectory are a combination of cohort and period effects. Of importance specifically for this study, is that except for the most recent birth cohorts, men seem to be on a higher earnings trajectory than women suggesting that the gender wage gap persists. At the same time, this difference seems to have been declining over time. A decomposition of earnings into age, cohort and period effects by gender will determine which effect has been most important in the decline of the gender wage gap over time.

Figure 31 plots the log of the raw gender wage gap by age and cohort. In figure 31a we use bracket weights to account for earnings given in brackets, while in figure 31b which seems to give smoother trajectories, these weights have not been applied.

(a) Bracket earnings corrected using bracket weights



(b) Bracket weights excluded

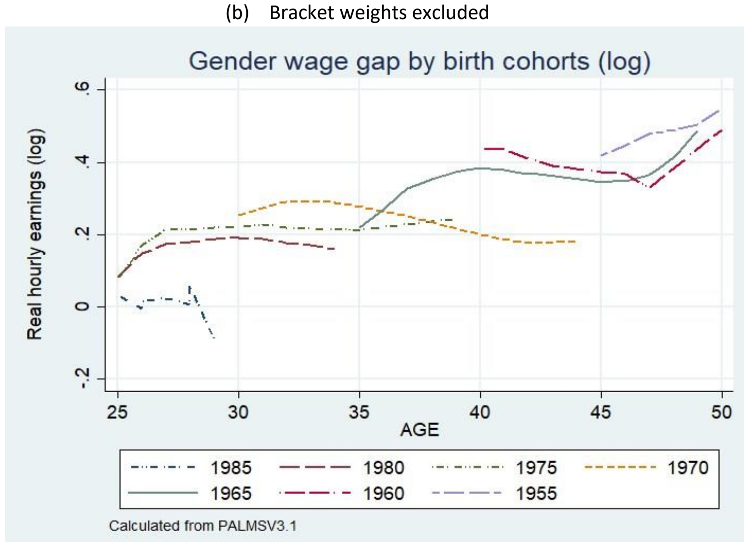


Figure 31: Life cycle trends of the gender wage gap

The figures confirm what we saw above in figure 30 and shows that the gap is smaller for more recent cohorts. This suggests that recent female cohorts have had more relative wage gains. Empirically, studies find that the gap increases with age (e.g. Bertrand et al. 2010; Goldin 2014). The explanation is that at the beginning of a career, wages are similar between men and women however, women are more likely to take time off due to family responsibilities therefore accumulating less hours of experience leading to a bigger wage gap as they age (Goldin 2014).

From figure 31 it is, however, difficult to see a specific pattern of the wage gap as cohorts age. It looks like for the 1965 and 1955 cohorts the gap seems to increase as the cohorts age whereas for the 1960 cohort the gap first declines and then increases after age 46. For the younger cohorts, the slopes of the wage gap profiles seem flatter implying that younger cohorts of women are less likely to have interruptions in their labour market participation. It is clear from the figure that more recent cohorts experience a lower wage gap. However, more recent cohorts are primarily observed early on in their career when the wage gap is expected to be lowest, and older cohorts are mainly observed later in their careers when the wage gap is declining. Therefore, what we see in figure 31 could be an age or a cohort effect or a combination of the two. To investigate this further, we formally decompose these effects using a multivariate regression.

4.6.3 Age, Period and Cohort Effects Decomposition Results

4.6.3.1 Female and Male Log Earnings

Age, cohort and period effects from the male and female log wage regressions as per equations (21) and (22) are plotted in figures 32, 33 and 34 (See table 16 in the appendix for the complete results). We include the intercepts[34] from the regressions to illustrate that men start at higher wages and that the wage gap persists even though female wages have been increasing over time and in some cases the male and female coefficients have converged.

Figure 32 shows that for both men and women, earnings increase with age. This result is significant for both men and women. Positive coefficients depict higher wages relative to the base category which is age 25 whereas a negative depicts lower wages relative to the base category. The figure shows that after age 30 women earn higher than men. However, when we include confidence bands around the estimates the result is not statistically significant (results not shown for clarity of the figure). We therefore urge

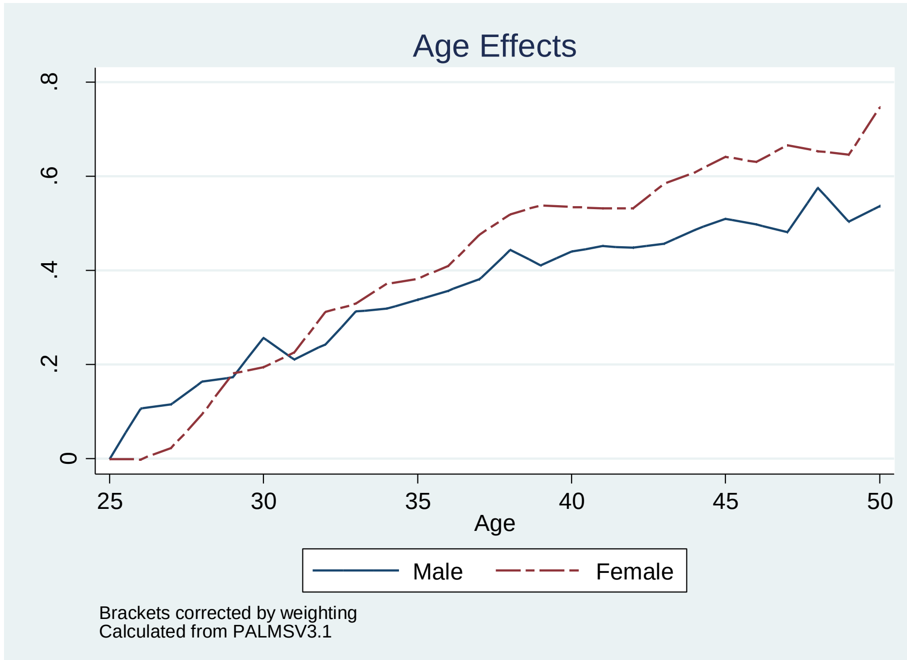
[34] The intercept from the female wage regression is log 1.133 while that from the male wage regression is log 1.553

caution in the interpretation of the age effects as the trend depicted by figure 32 could be as a result of selection effects. It could be the case that the women that drop out of the labour market are the low productivity women, leaving high productivity women in the labour market hence the higher returns to age. That the intercept from the male wage regression is higher and significant, however, indicates that the gender wage gap persists throughout the lifecycle.

What is also interesting about figure 32 is that, for both men and women, the age-earnings profile is contrary to the nonlinear relationship between age and earnings evident in the Mincerian wage regressions (Mincer 1974). It is expected that as workers age their productivity rises due to the accumulation of experience and on the job training. However, as they approach retirement, productivity will decline due to reduced investment on training and therefore earnings are expected to decline as well. What figure 32 implies is that for our sample, earnings are still on the rise even as individuals approach retirement. This is however not necessarily peculiar as decomposition studies from different countries find a wide range of age-earnings relationships, including steadily rising patterns for developing countries such as Taiwan (Smyk et al. 2014). However as indicated above, selection could be a problem in our dataset especially since descriptive analysis showed that there is a peculiar decline in employment and participation rates for both men and women after the age 35.

These results are contrary to findings by Grün (2003) who using three waves of OHS data and applying the Deaton normalization, found that wages increase till age 30 for both men and women and then start declining after that. More importantly for women she found a much steeper decline in wages as they aged, in that by age 40 women's wages were at a similar level to those just starting their careers. It is not clear what would account for the difference in results between our study and that of Grün (2003). The main difference is that Grün's study was for a different period (1995-1999) and she only used a limited number of surveys (OHS 1995, 1997 and 1999).

(a) Without the intercepts



(b) With intercepts

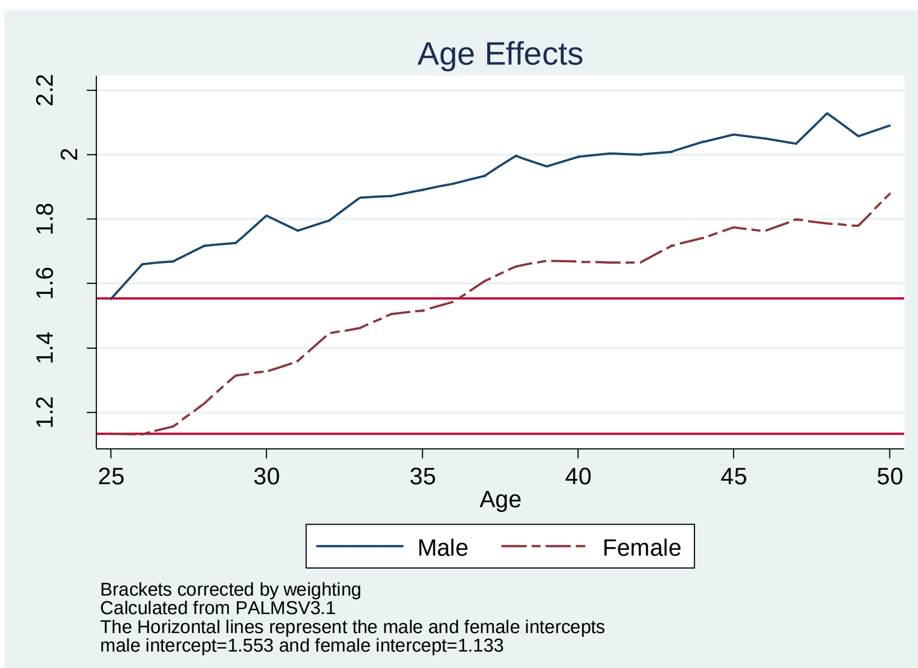
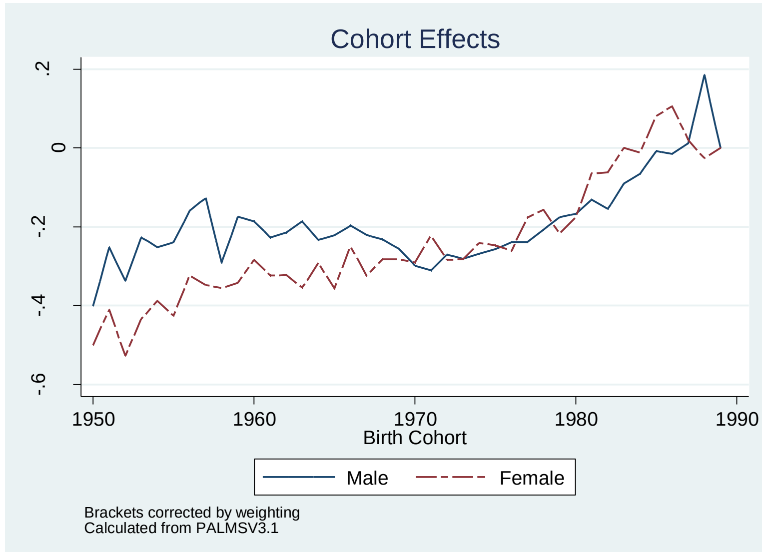


Figure 32: Coefficients on age dummies from the male and female wage regressions

(a) Cohort Effects without Intercept



(b) Cohort Effects with intercept

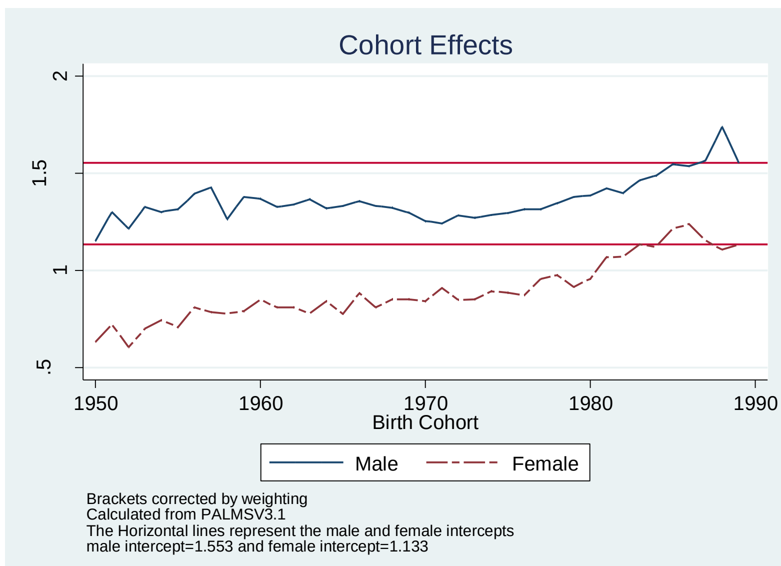


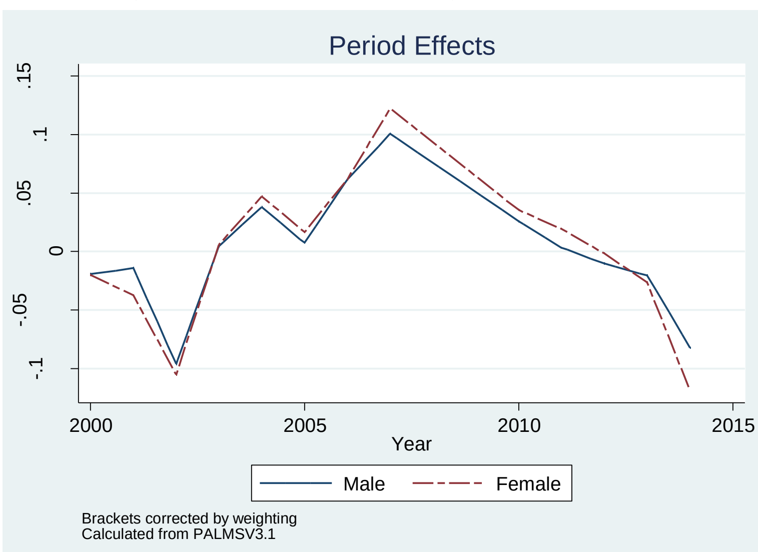
Figure 33: Cohort effects

Figure 33 plots the coefficients from the cohort dummies and shows that younger cohorts of men, those born between 1950 and 1970, experienced a decline in wages. For cohorts born after 1970 however, there was an increase in wages for both men and women, with the relative increase in wages higher for women. Given that we used the Deaton-Paxson normalization to decompose the age, period and cohort effects, it is possible and important to bear in mind that the trend evident above is partially attributable to period effects.

The trend in figure 33 could be linked to the demise of the apartheid regime and the introduction anti-discrimination legislation. The installation of a new democratic government may have opened up better paying occupations to women that were not previously accessible to them leading to the higher relative increase in wages for women compared to men. Additionally, Casale & Posel (2005) note that women may have become more "attractive" to employers for professional occupations due to affirmative action and equal opportunity programmes. The professional and occupational occupations which were previously dominated by the white sub-population group and men are on average higher paying therefore, that black women could now access them may have had an effect on the average earnings of women.

Moreover, black women may have benefited from the growth of the tertiary sectors including community services sector (Bhorat et al. 2014) this may have driven up female average wages leading to the decline of the gender wage gap. Additionally, the decline in employment in the primary sectors such as mining and agricultural sectors and the poor performance of the manufacturing sector in the period 2000-2012 (Bhorat et al. 2014) may have disadvantaged men more than women. This is because these sectors are male dominated therefore the decline in employment may have applied downward pressure on male wages. The decline of the gender wage gap evident in the cross-sectional analyses might therefore be due to a combination of deteriorating wages for men and higher relative wage gains for younger cohorts of women.

(a) Age, Cohort and Period Effects Decomposition: Period Effects



(b) Period Effects with intercept

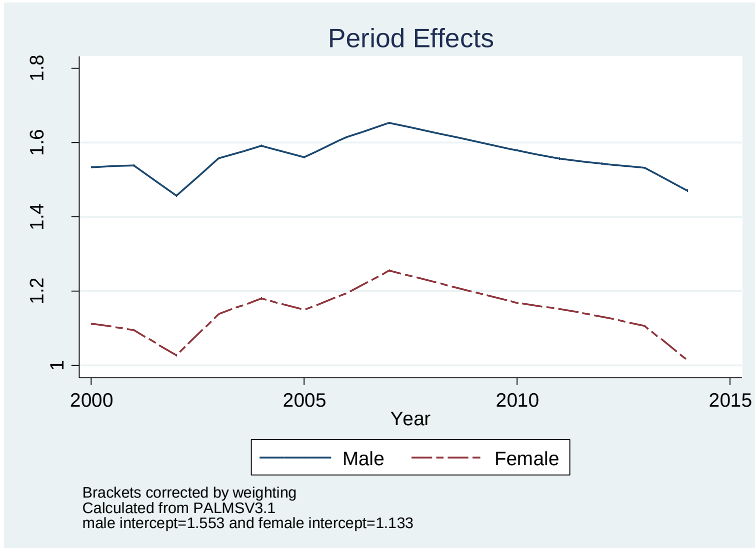


Figure 34: Period Effects

The period effects plotted in figure 34 show how wages change over time and, show an increase in wages between 2000 and 2007 and a decline afterwards. This result, similar to that of Burger & Von Fintel (2009)'s, shows that wages have a strong correlation with the economic cycle. This trend is consistent with the post-apartheid macro-economic environment which exhibited slow growth in the early 90s, a decline between 1997 and 2000 and a rise in economic growth after that (Makgetla 2004). The decline in wages after 2007 can partially be attributed to the 2007-2008 global financial crisis. Bhorat et al. (2014) note that the South African labour market is still trying to recover from the negative effects of the recession.

4.6.3.2 The Gender Wage Gap

Complete regression results for the gender wage gap are also presented in table 16 in the appendix. Positive coefficients mean that a variable contributes to the expansion of the gender wage gap while negative coefficients contribute to the narrowing of the gap. The results show that, as expected increased education contributes to the decline of the gender wage gap although the coefficients are not statistically significant. The results are also consistent with the trends we saw above for male and female wages.

Figure 35 is a graphical representation of age, cohort and period coefficients from equation (23). The period effects fluctuate between positive and negative values depicting a wage gap that is responsive to the macro-economic environment in South Africa discussed previously. There was a fall in the wage gap in the period 2000-2007, possibly a function of female wage growth during this period but this trend reversed after 2007.

Age effects depict an increase in the wage gap between age 25 and 45, then a decline after that. This is in line with findings in the literature that the gender wage gap increases during the child bearing years and declines later in life because some women especially the highly educated may take time off from the labour market and return later in life (Contreras et al. 2005; Polachek 1975a). Given the early departure from the labour market as seen in the labour force participation and employment trends for black men and women, it is possible that the decline of the gender wage gap after age 45 is due to a selection effect. It could be the case that the women that drop out of the labour market are low income earning individuals leaving either to take care of grandchildren or for other reasons (health) leaving high income earning women in the labour market and hence the decline in the wage gap.

The cohort effects figure shows that the wage gap has been declining systematically for younger cohorts. Younger cohorts of women have, on average, more education than the cohorts before them and men from the same cohort. These cohorts of women are more likely to be in a skilled profession compared to women born 30 years before them and who joined the labour market during the apartheid era. Marriage rates and trade union rates have greatly declined for recent cohorts. These changes over time have led to men and women in the labour market becoming more similar in terms of human capital characteristics. Additionally, in response to labour market transformation, recent cohorts of women may have become more attractive to employers as a previously disadvantaged group (Casale & Posel 2005) leading to more relative wage gains compared to men. All these factors combined, have in turn contributed to a narrowing gender wage gap at the mean over time. The decline of the gender wage gap evident in cross sectional analyses can therefore be linked to these more recent cohorts of women experiencing a lower gender wage gap a phenomenon referred to as ‘cohort replacement effect’.

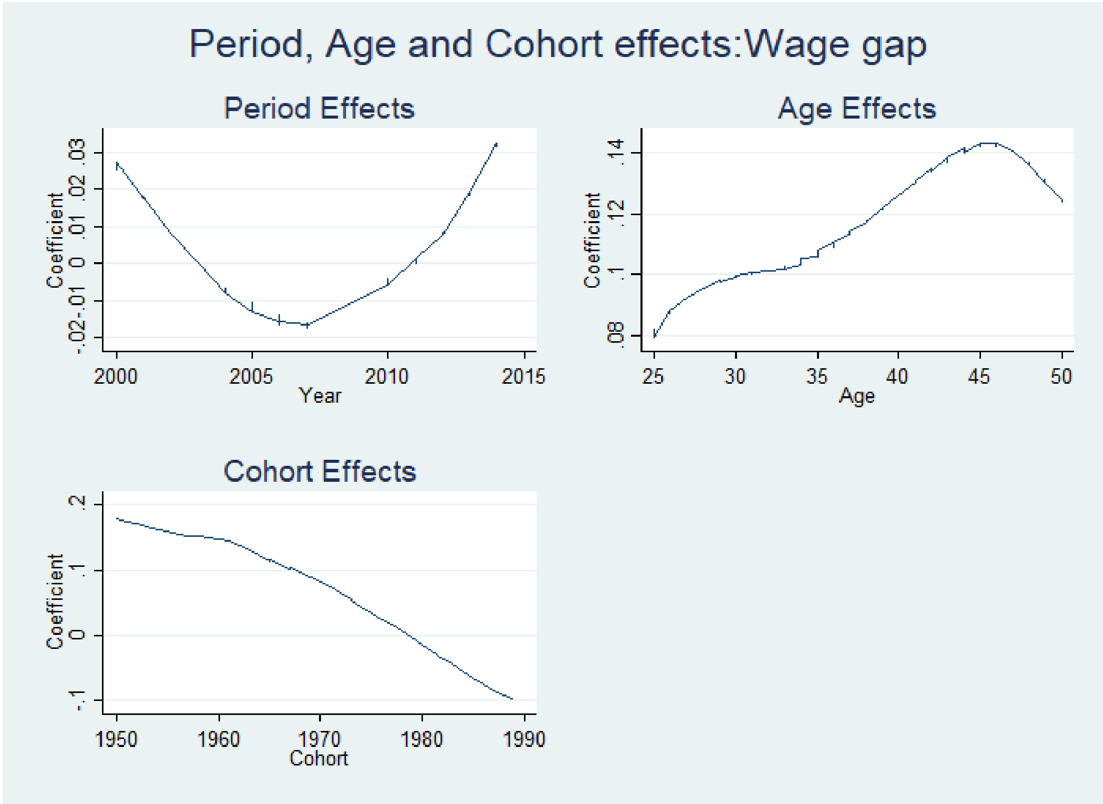


Figure 35: Age, period and cohort effects on the gender wage gap

Source: Own Calculation from PALMS V3.1

4.7 DISCUSSION

In this chapter, we analysed trends in real wages for African men and women in the South African labour market using synthetic cohort data. Life cycle trends in wages examined separately for men and women show that earnings increase with age and in addition, there is no statistical difference between the male and female returns to age (productivity). Results also show that the age-earnings profiles for African men and women do not display the expected 'hump shape' typical to Mincerian wage regressions. This is contradictory to the theory of human capital which assumes that productivity will decline after some point as individuals age and therefore wages should decline as well. One reason for the continual rise in earnings could be the fact that men and women are shown to be dropping out of the labour market earlier than the expected retirement age. We suspect health reasons could explain part of this early departure however this is an issue for further research. For women it could be that grandparents may be dropping out to look after grandchildren. This could be due to the fact that younger generations have on average higher education and therefore higher earning potential.

Age-cohort trends reveal that younger cohorts have experienced an increase in wages. Younger cohorts of men are however, not doing as well as women in terms of wage growth and for some cohorts a decline in wages was shown. This relative shift is part of the reason behind for the narrowing of the gender wage gap in recent years.

Results show that there have been generational effects on wages, schooling, marital status and union status. The youngest cohorts have on overage more education than their predecessors, are less likely to report being married and less likely to be in a trade union. We can link this to the narrowing of the gender wage gap for the youngest cohorts since previously, men were more likely to be in trade unions and this was associated with higher wages.

The recent decline of the gender wage gap as per cross sectional analysis, seems to be a result of a 'cohort replacement effect'. Where younger cohorts with better human capital characteristics have replaced older cohorts. This result is similar to Campbell & Pearlman (2013) who attribute the decline of the gender wage gap in the United States to cohort effects.

5 CONCLUSION

5.1 DATA QUALITY ISSUES

Unlike previous studies where measurement issues are not explicitly addressed, in this analysis we took a deeper look at data quality issues. Besides showing how results found in the literature have been affected by missing information and classification changes overtime, this thesis showed that there are several ways of dealing with breaks in the data.

We showed that utilising all data available can enable one to separate data related issues and real social changes. By applying both stochastic and multiple imputation methods we showed that there are ways of dealing with missing data. Additionally, constructing cohort data from repeated cross-sectional data is a way of connecting different data points to show a trend. In doing this, one is able to check if within cohort changes are plausible given that between cohort changes may be less affected by breaks than within cohort changes. In summary, we were able to address some data quality issues, but we conclude that data quality remains an issue of continuing research.

5.2 THE EVOLUTION OF THE GENDER WAGE GAP

This thesis set out to analyse the gender wage gap over the period 1993-2014. We examined the gender wage gap across the entire wage distribution and found that the gender wage gap has been stagnant at the median over the period analysed. This is interesting because with the introduction of anti-discrimination legislation and affirmative action, one would expect that the gender wage gap would have declined over time. This ties into other work on wages and wage inequality in South Africa. Wittenberg (2016a) finds that while the bottom part of the wage distribution has moved closer to the middle and the top part of the wage distribution has moved away from the middle, earnings at the median have been stagnant. He tries to locate the median worker in the economy and concludes that this worker is most likely African, male, with slightly more than 10 years of education and in their late thirties. In terms of occupation and industry, the median worker is most likely to be in the trade, manufacturing, services or finance industry doing elementary, service, crafts, or operational work. Our analysis of the trends in industry and occupation by gender showed that even though women have begun to join these occupations and industries, they are still male dominated. The gender wage gap can shift either due to changes in the labour market characteristics of individuals or due to a shift in the returns to those characteristics. Therefore, the fact that wages have stagnated, and the occupation and industry composition has not changed much, can partially explain the stagnation of the gender wage gap at the median.

Related to this, is the fact that from the demand side, the industry seems to be shifting towards employees with tertiary education. The returns to having less than 12 years of education have been declining over time for both men and women. While returns to completing 12 years of education (matric) are higher than not completing 12 years of education, they have also been declining over time. Returns to tertiary education on the other hand have been increasing. With minimum wage legislation helping boost wages in the bottom part of the distribution and the industry's demand for tertiary educated individuals pushing wages at the top end of the wage distribution up, it seems like the median worker has mainly been "forgotten" in the post-apartheid labour market and thus the stagnant median gender wage gap.

The size and reasons for the gender wage gap are heterogeneous across the wage distribution in the South African labour market. The experience of women at the top end of the wage distribution is different from the experience of women at the bottom or at the median. The implication is that policies aimed at narrowing the overall gender wage gap encounter different challenges and therefore require different solutions. While improved education and the implementation of minimum wage legislation has worked at the bottom end of the wage distribution, it will not work at the top end. We find that while better education contributes to reducing the unexplained wage gap at the bottom of the distribution, this is not the case at the top of the wage distribution.

A major reason for the persistent and increasing unexplained gender wage gap at the top end of the wage distribution is the heavy burden women shoulder due to lack of childcare facilities which affects their access to high paying positions and promotions. Policies to narrow the wage gap at the top end will need to focus on increasing the number of women in management and leadership roles. This will require interventions that make it possible for women to balance labour market and family responsibilities. That non-gender specific wage legislation i.e. the minimum wage legislation, worked to narrow the gender wage gap at the bottom of the wage distribution in South Africa, highlights that addressing the gender wage gap may require more than labour market legislation focused on gender equality. It may require a broader look at society and social norms and how these have or have not shifted over time.

From the cohort analysis, we find reason to believe that the decline of the gender wage gap at the mean is due to the youngest cohorts of women experiencing a lower gender wage gap due to more relative wage gains compared to men. We find reason to believe that continued political debate on labour market transformation after the demise of apartheid may have opened opportunities to women that were not accessible to them previously. Other than the fact that the youngest cohorts of women seem to have similar or better human capital characteristics than men, other factors such as labour market transformation may have made the recent cohorts of women relatively more employable in professional positions than their male counterparts thus reducing the gender wage gap at the mean.

That age-earnings profiles for men and women do not display the expected ‘hump shape’ from a Mincerian wage regression is an indication that wages do not decline with the decline in productivity (as workers age) or that those remaining in the labour market are selectively more productive than those that leave before their retirement age. We find reason to believe that selection after age 35 might be an issue in our data. Descriptive analysis revealed declines in employment and participation rates for both men and women after age 35. This issue requires further investigation. Future analysis should involve investigating the reasons for this early retirement for both men and women. A true panel data where the same individuals are followed over time would provide more answers. At the time of writing this thesis, the NIDS study had completed 4 waves of data collection. With more waves of the NIDS data it will be possible to carry out a panel analysis which will enrich results from this study.

REFERENCES

Abe, Yukiko. 2010. Equal Employment Opportunity Law and the gender wage gap in Japan: A cohort analysis. Journal of Asian Economics, 21(2), 142-155.

Aigner, Dennis J, & Cain, Glen G. 1977. Statistical theories of discrimination in labour markets. Industrial and Labour relations review, 30(1), 175-187.

Albrecht, James, Björklund, Anders, & Vroman, Susan. 2003. Is there a glass ceiling in Sweden? Journal of Labour economics, 21(1), 145-177.

Allison, Paul D. 2001. Missing Data: Sage University Papers Series on Quantitative Applications in the Social Sciences (07-136). Thousand Oaks, CA.

Altonji, Joseph G, & Blank, Rebecca M. 1999. Race and gender in the labour market. In: Ashenfelter, O., Card, D. (Eds.),. Pages 3143-3259 of: Handbook of Labour Economics, vol. 3C. Amsterdam, the Netherlands: Elsevier Science B.V.

Altonji, Joseph G, Bharadwaj, Prashant, & Lange, Fabian. 2012. Changes in the characteristics of American youth: Implications for adult outcomes. Journal of Labour Economics, 30(4), 783-828.

Angrist, Joshua D, & Pischke, Jörn-Steffen. 2008. Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Antecol, Heather, Jong, Anneke, & Steinberger, Michael. 2008. The sexual orientation wage gap: The role of occupational sorting and human capital. ILR Review, 61(4), 518-543.

Arrow, Kenneth Joseph. 1973. The Theory of Discrimination. In: Ashenfelter, O., and Rees, A.(Eds.),. Princeton, NJ: Princeton University Press.

Arulampalam, Wiji, Booth, Alison L, & Bryan, Mark L. 2007. Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. Industrial & Labour Relations Review, 60(2), 163-186.

Atkinson, Anthony B, & Bourguignon, François. 2014. Handbook of Income Distribution. Vol. 2A-2B. Elsevier.

Attanasio, Orazio P. 1993. A cohort analysis of saving behaviour by US households. Tech. rept. National Bureau of Economic Research.

Baltagi, Badi. 2008. Econometric analysis of panel data. John Wiley & Sons.

Baron, Juan D, & Cobb-Clark, Deborah A. 2010. Occupational segregation and the gender wage gap in private-and public-sector employment: a distributional analysis. Economic Record, 86(273), 227-246.

Barsky, Robert, Bound, John, Charles, Kerwin Ko', & Lupton, Joseph P. 2002. Accounting for the black white wealth gap: a nonparametric approach. Journal of the American Statistical Association, 97(459), 663-673.

Beaudry, Paul, & Lemieux, Thomas. 1999. Evolution of the female labour force participation rate in Canada, 1976-1994: A cohort analysis. Canadian Business Economics, 7(2), 57.

Becker, Gary S. 1971. Economics of Discrimination. 2nd edn. Chicago: University of Chicago Press.

Bergmann, Barbara R. 1974. Occupational segregation, wages and profits when employers discriminate by race or sex. Eastern Economic Journal, 1(2), 103-110.

Bertrand, Marianne, Goldin, Claudia, & Katz, Lawrence F. 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. American Economic Journal: Applied Economics, 2(3), 228-255.

Bhorat, Haroon. 2004. Labour Market Challenges in the Post-Apartheid South Africa. South African Journal of Economics, 72(5), 940-977.

Bhorat, Haroon, & Goga, Sumayya. 2013. The gender wage gap in post-apartheid South Africa: A re-examination. Journal of African Economies, 22(2), 1-22.

Bhorat, Haroon, Goga, Sumayya, & Stanwix, Benjamin. 2014b. Skills-biased labour demand and the pursuit of inclusive growth in South Africa. Tech. rept. WIDER Working Paper.

Bhorat, Haroon, Hirsch, Alan, Kanbur, Ravi, & Ncube, Mthuli. 2014. The oxford companion to the economics of South Africa. OUP Oxford.

Bhorat, Haroon, Kanbur, Ravi, & Mayet, Natasha. 2012. Minimum wage violation in South Africa. International Labour Review, 151(3), 277-287.

Bhorat, Haroon, Kanbur, Ravi, & Mayet, Natasha. 2013. The impact of sectoral minimum wage laws on employment, wages, and hours of work in South Africa. IZA Journal of Labour & Development, 2(1), 1.

Bhorat, H, & Rooney, C. 2017. State of Manufacturing in SA. Development Policy Research Unit Working Paper 201702.

Blau, Francine D, & Kahn, Lawrence M. 1997. Swimming upstream: Trends in the gender wage differential in the 1980s. Journal of labour Economics, 15(1, Part 1), 1-42.

Blau, Francine D, & Kahn, Lawrence M. 2000. Gender differences in pay. Tech. rept. National bureau of economic research.

Blau, Francine D, & Kahn, Lawrence M. 2006. The US gender pay gap in the 1990s: Slowing convergence. Industrial & Labor Relations Review, 60(1), 45-66.

Blau, Francine D, & Kahn, Lawrence M. 2016. The Gender Wage Gap: Extent, Trends, and Explanations. IZA Discussion Paper, No.9656.

Blinder, Alan S. 1973. Wage discrimination: reduced form and structural estimates. Journal of Human resources, 8(4), 436 455. Branson, Nicola, & Leibbrandt, Murray. 2013. Educational attainment and labour market outcomes in South Africa, 1994-2010.

Branson, Nicola, & Wittenberg, Martin. 2007. The measurement of employment status in South Africa using cohort analysis, 1994-2004. South African Journal of Economics, 75(2), 313-326.

Branson, Nicola, & Wittenberg, Martin. 2014. Reweighting South African national household survey data to create a consistent series over time: A cross-entropy estimation approach. South African Journal of Economics, 82(1), 19-38.

Branson, Nicola, Ardington, Cally, Lam, David, & Leibbrandt, Murray. 2013. Changes in education, employment and earnings in South Africa: A cohort analysis. Southern Africa Labour and Development Research Unit Working Paper Number 105.

Browning, Martin, & Lusardi, Annamaria. 1996. Household saving: Micro theories and micro facts. Journal of Economic literature, 34(4), 1797-1855.

Budlender, Debbie. 2010. Decent work for domestic workers. Study prepared for the SERVICES Sector Education and Training Authority. Johannesburg: Community Agency for Social Enquiry (CASE).

Budlender, Debbie. 2011. Measuring the economic and social value of domestic work: Conceptual and methodological framework. ILO.

Budlender, Debbie. 2016. The introduction of a minimum wage for domestic workers in South Africa. Conditions of work and employment series; No. 72. International Labour Office, Inclusive Labour Markets, Labour Relations and Working Conditions Branch., Geneva: ILO.

Burger, Rulof, & Jafta, Rachel. 2010. Affirmative action in South Africa: an empirical assessment of the impact on labour market outcomes. CRISE (Centre for Research on Inequality, Human Security and Ethnicity) Working Paper, 76, 09-36.

Burger, Rulof, & Von Fintel, Dieter. 2009. Determining the causes of the rising South African unemployment rate: An age, period and generational analysis. Economic Research Southern Africa (ERSA) Working Papers.

Burger, Rulof, & Woolard, Ingrid. 2005. The state of the labour market in South Africa after the first decade of democracy. Journal of vocational education and training, 57(4), 453-476.

Burger, Rulof, & Yu, Derek. 2007. Wage trends in post-apartheid South Africa: Constructing an earnings series from household survey data. Working Paper 07/117. University of Cape Town: DPRU Working Paper.

Butcher, Kristin F, & Rouse, Cecilia Elena. 2001. Wage effects of unions and industrial councils in South Africa. ILR Review, 54(2), 349-374.

Campbell, Colin, & Pearlman, Jessica. 2013. Period effects, cohort effects, and the narrowing gender wage gap. Social science research, 42(6), 1693-1711.

Casale, Daniela. 2004. What has the feminisation of the labour market bought women in South Africa? Trends in labour force participation, employment and earnings, 1995-2001. Journal of Interdisciplinary Economics, 15(3-4), 251-275.

Casale, Daniela, Muller, Colette, & Posel, Dorrit. 2004. Two million net new jobs a reconsideration of the rise in employment in South Africa, 1995-2003. South African Journal of Economics, 72(5), 978-1002.

Casale, Daniela, & Posel, Dorrit. 2002. The continued feminisation of the labour force in South Africa. South African Journal of Economics, 70(1), 156-184.

Casale, Daniela, & Posel, Dorrit. 2005. Women and the economy: how far have we come? Agenda, 19(64), 21-29.

Casale, Daniela, & Posel, Dorrit. 2010. The male marital earnings premium in the context of bridewealth payments: Evidence from South Africa. Economic Development and Cultural Change, 58(2), 211-230.

Chi, Wei, & Li, Bo. 2008. Glass ceiling or sticky floor? Examining the gender earnings differential across the earnings distribution in urban China, 1987-2004. Journal of Comparative Economics, 36(2), 243-263.

Chung, YoonKyung, Downs, Barbara, Sandler, Danielle H, Sienkiewicz, Robert, et al. 2017. The Parental Gender Earnings Gap in the United States. Tech. rept.

Clark, Andrew, & D'Ambrosio, Conchita. 2015. Attitudes to Income Inequality: Experimental and Survey Evidence. In: Handbook of Income Distribution Volume 2A. Amsterdam: Elsevier.

Cleveland, William S. 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American statistical association, 74(368), 829-836.

Cock, Jacklyn. 1980. Maids & madams: a study in the politics of exploitation. Ravan Press.

Contreras, Dante, Puentes, Esteban, & Bravo, David. 2005. Female labour force participation in greater Santiago, Chile: 1957-1997. A synthetic cohort analysis. Journal of International Development, 17(2), 169-186.

Coate, Stephen, & Loury, Glenn C. 1993. Will affirmative-action policies eliminate negative stereotypes? The American Economic Review, 83(5), 1220-1240.

Deaton, Angus. 1985. Panel data from time series of cross-sections. Journal of econometrics, 30(1), 109-126.

Deaton, Angus. 1997. The analysis of household surveys: a micro econometric approach to development policy. Baltimore: Johns Hopkins University Press.

Deaton, Angus S, & Paxson, Christina. 1994. Saving, growth, and aging in Taiwan. Pages 331-362 of: Studies in the Economics of Aging. University of Chicago Press.

Devereux, Paul J. 2007. Small-sample bias in synthetic cohort models of labour supply. Journal of Applied Econometrics, 22(4), 839-848.

DiNardo, John, Fortin, Nicole, & Lemieux, Thomas. 1996. Labour Market Institutions and the Distribution of Wages, 1973-1992: A Semi-Parametric Approach. Econometrica, 64(5), 1001-1044.

Dinkelman, Taryn, & Ranchhod, Vimal. 2010. Evidence on the impact of minimum wage laws in an informal sector: Domestic workers in South Africa. A Southern Africa Labour and Development Research Unit Working Paper Number 44. Cape Town: SALDRU, University of Cape Town.

Drolet, Marie. 2011. Why has the gender wage gap narrowed? Perspectives on Labour and Income, 23(1), 5.

Easterlin, Richard A. 1968. The American baby boom in historical perspective. Pages 77 110 of: Population, Labour Force, and Long Swings in Economic Growth: The American Experience. NBER.

Fedderke, Johannes. 2010. Sustainable growth in South Africa, in Luiz de Mello (ed.). Sustainability in Brazil, China, India, Indonesia and South Africa, OECD Publishing, Paris.

Firpo, Sergio, Fortin, Nicole M, & Lemieux, Thomas. 2009. Unconditional quantile regressions. Econometrica, 77(3), 953-973.

Fitzenberger, Bernd, & Wunderlich, Gaby. 2002. Gender wage differences in West Germany: a cohort analysis. German Economic Review, 3(4), 379-414.

Fortin, Nicole M. 2008. The gender wage gap among young adults in the United States the importance of money versus people. Journal of Human Resources, 43(4), 884-918.

Fortin, Nicole, Lemieux, Thomas, & Firpo, Sergio. 2011. "Decomposition methods in economics," in O.Ashenfelter and D.Card, eds.,. Handbook of Economics, 4, 1-102.

Gaitskell, Deborah, Kimble, Judy, Maconachie, Moira, & Unterhalter, Elaine. 1983. Class, race and gender: Domestic workers in South Africa. Review of African Political Economy, 10(27-28), 86-108.

Ganguli, Ina, Hausmann, Ricardo, & Viarengo, Martina. 2014. Closing the gender gap in education: What is the state of gaps in labour force participation for women, wives and mothers? International Labour Review, 153(2), 173-207.

Gardeazabal, Javier, & Ugidos, Arantza. 2004. More on identification in detailed wage decompositions. Review of Economics and Statistics, 86(4), 1034-1036.

Gardes, François, Duncan, Greg J, Gaubert, Patrice, Gurgand, Marc, & Starzec, Christophe. 2005. Panel and pseudo-panel estimation of cross-sectional and time series elasticities of food consumption: The case of US and Polish data. Journal of Business & Economic Statistics, 23(2), 242-253.

Gelb, Stephen. 2003. Inequality in South Africa: Nature, causes and responses. Edge Institute Johannesburg.

Goldin, Claudia. 2014. A grand gender convergence: Its last chapter. The American Economic Review, 104(4), 1091-1119.

Goldin, Claudia, & Katz, Lawrence F. 2002. The power of the pill: oral contraceptives and women's career and marriage decisions. Journal of Political Economy, 110(4), 730-770.

Goraus, Karolina, Tyrowicz, Joanna, & Velde, Lucas. 2015. Which Gender Wage Gap Estimates to Trust? A Comparative Analysis. Review of Income and Wealth.

Gradín, Carlos. 2018. Occupational gender segregation in post-apartheid South Africa. WIDER Working Paper number 53.

Grün, Carola. 2003. Racial and gender wage differentials in South Africa: What can cohort data tell? Tech. rept. Munich Discussion Paper.

Grün, Carola. 2004. Direct and indirect gender discrimination in the South African labour market. International Journal of Manpower, 25(3/4), 321-342.

Heap, Anthea. 2008. Earnings Inequality in South Africa: Decomposing Changes between 1995 and 2006. M.Phil. thesis.

Heckman, James. 1979. Sample Selection Bias as a Specification Error. Econometrics, 47(1), 153-161.

Heckman, James, & Robb, Richard. 1985. Using longitudinal data to estimate age, period and cohort effects in earnings equations. Pages 137-150 of: Cohort analysis in social research. Springer.

Heintz, James, & Posel, Dorrit. 2008. Revisiting informal employment and segmentation in the South African labour market. South African Journal of Economics, 76(1), 26-44.

Hertz, Tom. 2005. The effect of minimum wages on the employment and earnings of South Africa's domestic service workers. Development Policy Research Unit Working Paper 05/99. Cape Town: University of Cape Town.

Hinks, Timothy. 2002. Gender wage differentials and discrimination in the New South Africa. Applied Economics, 34(16), 2043-2052.

Hirano, Keisuke, Imbens, Guido W, & Ridder, Geert. 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica, 71(4), 1161-1189.

Hirsch, Boris. 2016. Gender wage discrimination. IZA World of Labour.

International Labour Organisation. 2013. Domestic workers across the world: Global and regional statistics and the extent of legal protection. International Labour Office Geneva.

Jann, Ben. 2008. The Blinder-Oaxaca decomposition for linear regression models. The Stata Journal, 8(4), 453-479.

Jones, Frank L, & Kelley, Jonathan. 1984. Decomposing differences between groups a cautionary note on measuring discrimination. Sociological Methods & Research, 12(3), 323-343.

Juhn, Chinhui, Murphy, Kevin M, & Pierce, Brooks. 1993. Wage inequality and the rise in returns to skill. Journal of political Economy, 410-442.

Kassenboehmer, Sonja C, & Sinning, Mathias G. 2014. Distributional changes in the gender wage gap. Industrial & Labour Relations Review, 67(2), 335-361.

Kee, Hiau Joo. 2006. Glass ceiling or sticky floor? Exploring the Australian gender pay gap. Economic Record, 82(259), 408-427.

Kerr, A, Lam, D, & Wittenberg, M. 2013. Post-Apartheid Labour Market Series. [dataset]. Version 2.1. Cape Town: DataFirst [data producer and distributor], 2013.

Kerr, A, Lam, D, & Wittenberg, M. 2016. Post-Apartheid Labour Market Series. [dataset]. Version 3.1. Cape Town: DataFirst [data producer and distributor], 2016.

Kerr, Andrew, & Wittenberg, Martin. 2015. Sampling methodology and fieldwork changes in the October Household Surveys and Labour Force Surveys. Development Southern Africa, 32(5), 603-612.

Kerr, Andrew, & Wittenberg, Martin. 2016. Public sector wages and employment in South Africa. Unpublished working paper, REDI.

Keswell, Malcolm. 2004. Education and Racial Inequality in Post-Apartheid South Africa. Working Paper No. 2004-02-008, Santa Fe Institute, Santa Fe, New Mexico.

Keswell, Malcolm, & Poswell, Laura. 2004. Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. South African Journal of Economics, 72(4), 834-860.

Kumchulesi, Grace. 2010. An economic analysis of declining marriages in post-apartheid South Africa, 1995-2006. Ph.D. thesis, University of Cape Town.

Landis, Elizabeth S. 1975. African Women under Apartheid. Africa Fund.

Lang, Kevin, & Lehmann, Jee-Yeon K. 2011. Racial discrimination in the labour market: Theory and empirics. Tech. rept. National Bureau of Economic Research.

Leibbrandt, Murray, Woolard, Ingrid, Finn, Arden, & Argent, Jonathan. 2010b. Trends in South African income distribution and poverty since the fall of apartheid. OECD Social, Employment and Migration Working Papers, No. 101. OECD Publishing.

Leibbrandt, Murray, Woolard, Ingrid, McEwen, Hayley, & Koep, Charlotte. 2010a. Employment and inequality outcomes in South Africa. University of Cape Town: Southern Africa Labour and Development Research Unit.

Lund, Frances, & Budlender, Debbie. 2009. Paid Care Providers in South Africa: Nurses, Domestic Workers and Home-Based Care Workers. Project on the Political and Social Economy of Care, UNRISD, Geneva.

Luo, Liying. 2013. Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. Demography, 50(6), 1945-1967.

Machado, Jose AF, & Mata, Jose. 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. Journal of applied Econometrics, 20(4), 445-465.

Machemedze, T, Kerr, A, & Wittenberg, M. 2014. Recalibrating the OHSs to adjust for sampling changes. A DataFirst Technical Paper 28. Cape Town: DataFirst, University of Cape Town.

Makgetla, Neva Seidman. 2004. The post-apartheid economy. Review of African Political Economy, 31(100), 263-281.

Manning, Alan, & Swaffield, Joanna. 2008. The gender gap in early-career wage growth. The Economic Journal, 118(530), 983-1024.

Mason, William M, & Fienberg, Stephen. 2012. Cohort analysis in social research: Beyond the identification problem. Springer Science & Business Media.

Meng, Yang, Brennan, Alan, Purshouse, Robin, Hill-McManus, Daniel, Angus, Colin, Holmes, John, & Meier, Petra Sylvia. 2014. Estimation of own and cross price elasticities of alcohol demand in the UK a pseudo-panel approach using the Living Costs and Food Survey 2001-2009. Journal of health economics, 34, 96-103.

Mincer, Jacob A. 1974. Age and Experience Pro les of earnings. In Schooling, experience, and earnings. (pp. 64-82). NBER.

Mincer, Jacob, & Polacheck, Solomon. 1974. Family investments in human capital: Earnings of women. Pages 397-431 of: Economics of the Family: Marriage, Children, and Human Capital. University of Chicago Press.

Muller, Colette. 2009. Trends in the gender wage gap and gender discrimination among part-time and fulltime workers in post-apartheid South Africa. Working paper number 124. Durban, South Africa: University of KwaZulu Natal.

Nattrass, Nicoli, & Seekings, Jeremy. 2001. "Two nations"? Race and economic inequality in South Africa today. Daedalus, 130(1), 45-70.

Neumark, David. 1988. Employers' discriminatory behaviour and the estimation of wage discrimination. Journal of Human resources, 23(3), 279-295.

Neyens, E, & Wittenberg, Martin. 2016. Changes in Self-Employment in the Agricultural Sector, South Africa: 1994-2012. Working Paper, DataFirst University of Cape Town.

Ntuli, Miracle. 2007a. Determinants of South African women's labour force participation, 1995-2004. IZA Discussion Paper.

Ntuli, Miracle. 2007b. Exploring Gender Wage Discrimination in South Africa, 1995-2004: A Quantile Regression Approach. IPC Working paper series No.56. Michigan: Gerald R.Ford School of public policy, University of Michigan.

Ntuli, Miracle, & Wittenberg, Martin. 2013. Determinants of black women's labour force participation in post-apartheid South Africa. Journal of African economies, 22(3), 347-374.

Oaxaca, Ronald. 1973. Male-female wage differentials in urban labour markets. International economic review, 14(3), 693-709.

Oaxaca, Ronald L. 2007. The challenge of measuring labour market discrimination against women. Swedish Economic Policy Review, 14(1), 199.

Oaxaca, Ronald L, & Ransom, Michael R. 1994. On discrimination and the decomposition of wage differentials. Journal of econometrics, 61(1), 5-21.

Oaxaca, Ronald L, & Ransom, Michael R. 1999. Identification in detailed wage decompositions. Review of Economics and Statistics, 81(1), 154-157.

Okojie, Christiana, & Shimeles, Abebe. 2006. Inequality in sub-Saharan Africa. The Inter-Regional Inequality Facility. Overseas Development Institute (ODI). London.

O'Neill, June E, & O'Neill, Dave M. 2006. What do wage differentials tell about labour market discrimination? Pages 293-357 of: The economics of immigration and social diversity. Emerald Group Publishing Limited.

Petrongolo, Barbara, & Olivetti, Claudia. 2006. Unequal pay or unequal employment? A cross-country analysis of gender gaps.

Polachek, Solomon W. 1975a. Discontinuous labour force participation and its effect on women's market earnings. Sex, discrimination, and the division of labour, 90-122.

Polachek, Solomon W. 1975b. Differences in expected post-school investment as a determinant of market wage differentials. International Economic Review, 451-470.

Polachek, Solomon W. 2006. How the Human Capital Model Explains Why the Gender Wage Gap Narrowed. In: The Declining Significance of Gender? Francine D. Blau, Mary C. Brinton, and David B. Grusky, (Eds.),. 102-124.

Polachek, Solomon W. 2014. Equal pay legislation and the gender wage gap. IZA World of Labour.

Posel, Dorrit. 2004. Have migration patterns in post-apartheid South Africa changed? Journal of Interdisciplinary Economics, 15(3-4), 277-292.

Posel, Dorrit. 2014. Gender Inequality, in H. Bhorat, A. Hirsch, R. Kanbur and M.Ncube (eds). Oxford Companion to the economics of South Africa, Oxford University Press, Oxford, 303-310.

Posel, Dorrit, & Casale, Daniela. 2001. Gender aggregates: Women subsistence farmers affect the unemployment count. Agenda, 49, 82-88.

Posel, Dorrit, & Casale, Daniela. 2006. Who Replies in Brackets and what are the Implications for Earnings Estimates?: An Analysis of Earnings Data from South Africa. Economic Research Southern Africa (ERSA) Working Paper 7.

Posel, Dorrit, & Muller, Colette. 2008. Is There Evidence Of a Wage Penalty to Female Part-Time Employment In South Africa? South African Journal of Economics, 76(3), 466-479.

Posel, Dorrit, & Rogan, Michael. 2000. Gendered trends in income poverty in post-apartheid South Africa, 1997-2006. Ten Years of War against Poverty: What Have We Learned Since, 2010-2020.

Puhani, Patrick. 2000. The Heckman correction for sample selection and its critique. Journal of economic surveys, 14(1), 53-68.

Ranchhod, V. 2010. Labour force participation and employment in South Africa: evidence from wave 1 of the National Income Dynamics Study. Studies in Economics and Econometrics, 34(3), 111-127.

Rospabé, Sandrine. 2001. An empirical evaluation of gender discrimination in employment, occupation attainment and wage in South Africa in the late 1990s. In: DPRU/FES workshop on Labour markets and poverty in South Africa, Johannesburg, November. Citeseer.

Rubin, Donald B. 1996. Multiple imputation after 18+ years. Journal of the American statistical Association, 91(434), 473-489.

Russell, Jeffrey E, & Fraas, John W. 2005. An application of panel regression to pseudo panel data. Multiple linear regression viewpoints, 31(1), 1-15.

Schultz, T Paul, & Mwabu, Germano. 1998. Labour unions and the distribution of wages and employment in South Africa. ILR Review, 51(4), 680-703.

Seekings, Jeremy, & Nattrass, Nicoli. 2008. Class, race, and inequality in South Africa. Yale University Press.

Shepherd, Debra. 2008. Post-apartheid trends in gender discrimination in South Africa: Analysis through decomposition techniques. Stellenbosch, University of Stellenbosch. Retrieved October, 5, 2011.

Smyk, Magdalena, Tyrowicz, Joanna, & Liberda, Barbara. 2014. Age productivity patterns in talent occupations for men and women: a decomposition. Post-Communist Economies, 26(3), 401-414.

Sulla, Victor, & Zikhali, Precious. 2018. Overcoming Poverty and Inequality in South Africa: An Assessment of Drivers, Constraints and Opportunities (English). Washington, DC: World Bank.

Thornton, Robert J., Rodgers, James D., & Brookshire, Michael L. 1997. On the interpretation of age-earnings pro les. Journal of Labour Research, 18(2), 351-365.

Tyrowicz, Joanna, van der Velde, Lucas, van Staveren, Irene, et al. 2015. Differences in the Estimates of Gender Wage Gap over the Life Cycle. Tech. rept.

Van Klaveren, Maarten, Tijdens, Kea, Hughie-Williams, Melanie, & Martin, N Ramos. 2009. An overview of women's work and employment in South Africa. Amsterdam: AIAS Working Paper, 09-79.

Verbeek, Marno, & Nijman, Theo. 1993. Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections. Journal of Econometrics, 59(1-2), 125-136.

Warunsiri, Sasiwimon, & McNown, Robert. 2010. The returns to education in Thailand: A pseudo-panel approach. World Development, 38(11), 1616-1625.

Weichselbaumer, Doris, & Winter-Ebmer, Rudolf. 2005. A meta-analysis of the international gender wage gap. Journal of Economic Surveys, 19(3), 479-511.

Welch, Finis. 1979. Effects of cohort size on earnings: the baby boom babies' financial bust. Journal of Political Economy, 87(5, Part 2), S65-S97.

White, Ian R, Royston, Patrick, & Wood, Angela M. 2011. Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine, 30(4), 377-399.

Winter, Carolyn. 1999. Women Workers in South Africa: Participation, Pay and Prejudice in the Formal Labour Market. World Bank Departmental Working Paper, No.19752, Washington, D.C.

Wittenberg, M. 2014. Data issues in South Africa, in H. Bhorat, A. Hirsch, R. Kanbur and M.Ncube (eds). Oxford Companion to the economics of South Africa, Oxford University Press, Oxford, 79-83.

Wittenberg, Martin. 2016a. Trends in Earnings and Earnings Inequality in South Africa: 1993-2014. Southern Africa Labour and Development Research Unit.

Wittenberg, Martin. 2016b. Wages and Wage Inequality in South Africa 1994-2011: Part 1 Wage Measurement and Trends. South African Journal of Economics.

Wittenberg, Martin, & Pirouz, Farah. 2013 (September). The measurement of earnings in the post-Apartheid period: An overview. Technical paper 23. DataFirst.

Wodon, Quentin T, & de la Brière, Bénédicte. 2018. Unrealized Potential: The High Cost of Gender Inequality in Earnings. The Cost of Gender Inequality. Washington, DC: World Bank.

Yang, Yang, & Land, Kenneth C. 2006. A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. Sociological methodology, 36(1), 75-97.

Yang, Yang, Schulhofer-Wohl, Sam, Fu, Wenjiang J, & Land, Kenneth C. 2008. The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It. American Journal of Sociology, 113(6), 1697-1736.

Yu, Derek. 2007. The comparability of the Statistics South Africa October Household Surveys and Labour Force Surveys. Stellenbosch Economic Working Papers 17/07. Stellenbosch University, Stellenbosch.

Yu, Derek. 2008. The South African Labour Market: 1995-2006.

Yu, Derek. 2009. The comparability of Labour Force Survey (LFS) and Quarterly Labour Force Survey (QLFS). Tech. rept.

Yun, Myeong-Su. 2005. A simple solution to the identification problem in detailed wage decompositions. Economic inquiry, 43(4), 766-772.

APPENDIX A: CHAPTER THREE

A1 Descriptive Statistics for wage employed women versus unemployed women

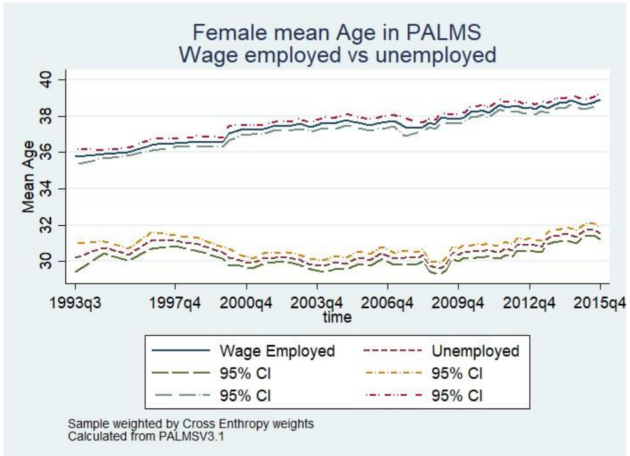


Figure 36: Mean Age: Wage employed vs Unemployed

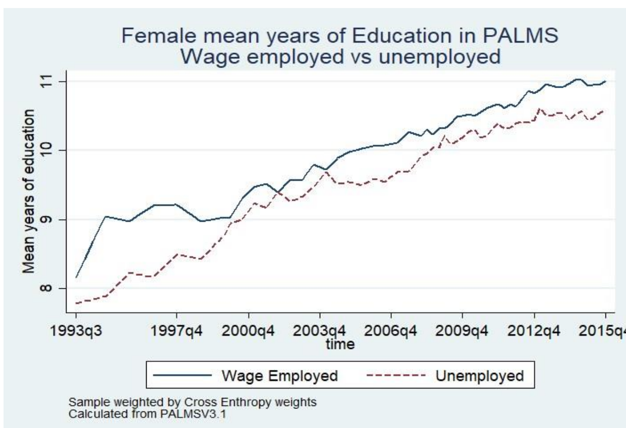


Figure 37: Years of education: Wage employed vs Unemployed

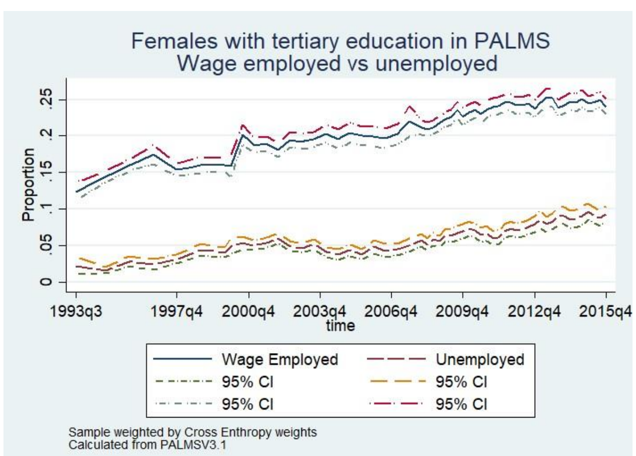


Figure 38: Tertiary education: Wage employed vs Unemployed

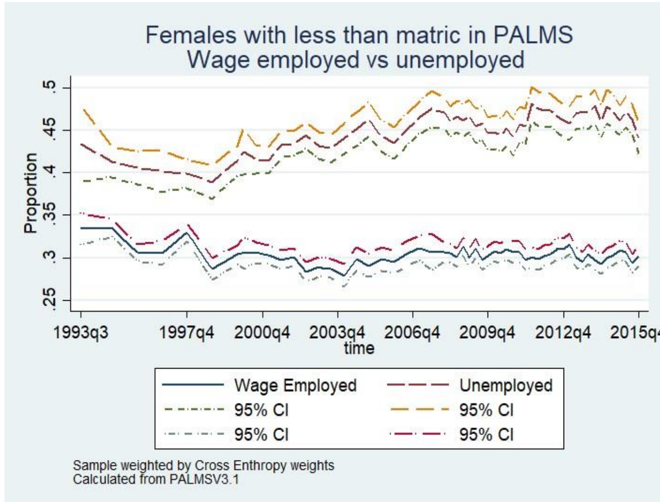


Figure 39: Less than Matric: Wage employed vs Unemployed

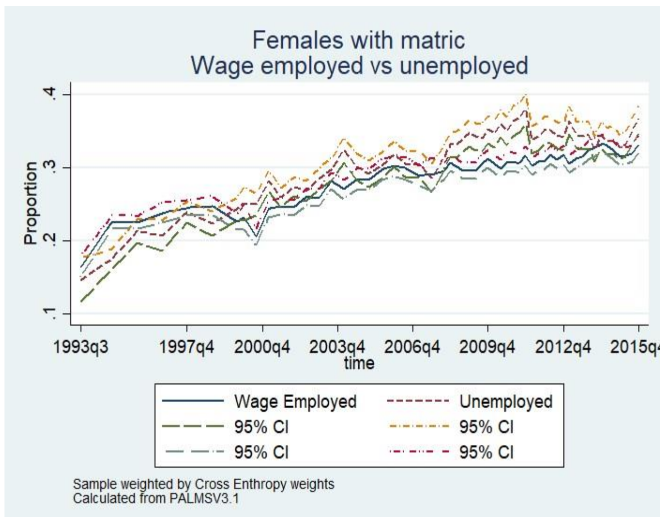


Figure 40: Matric: Wage employed vs Unemployed

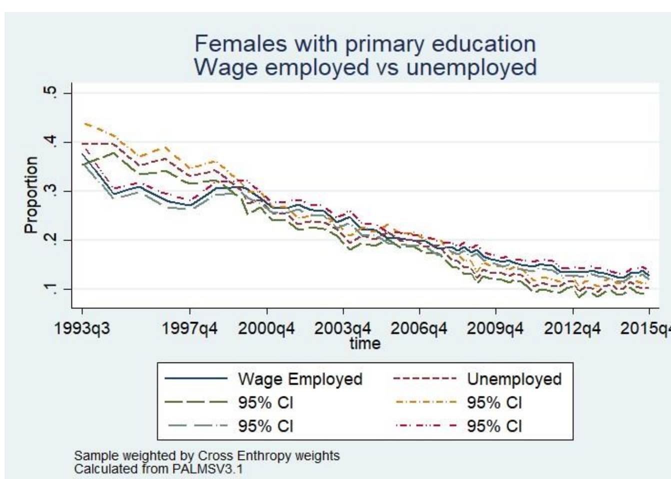


Figure 41: Primary education: Wage employed vs Unemployed

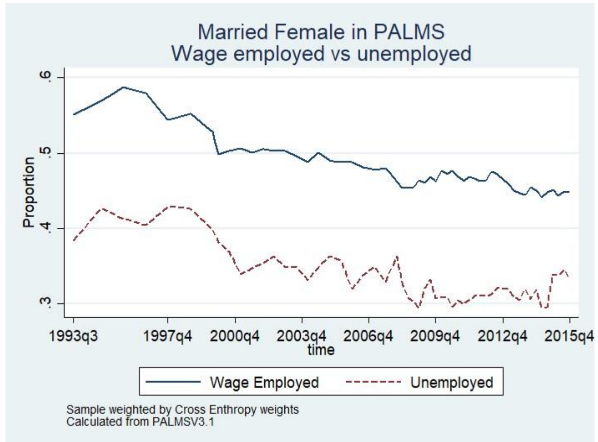


Figure 42: Proportion Married: Wage employed vs Unemployed

Descriptive Statistics by gender: Female vs Male

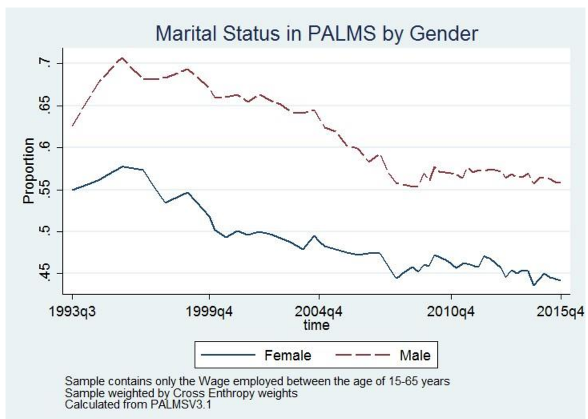


Figure 43: Proportion Married by gender

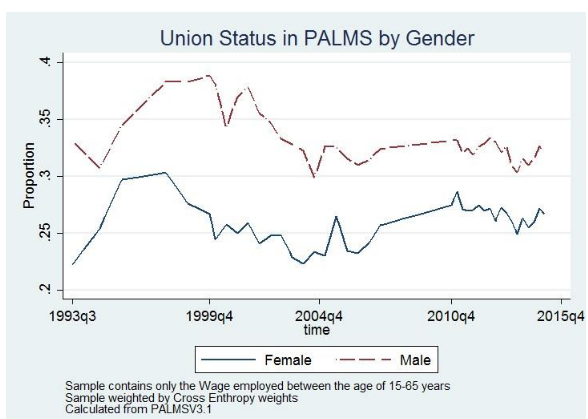


Figure 44: Proportion in a trade union by gender

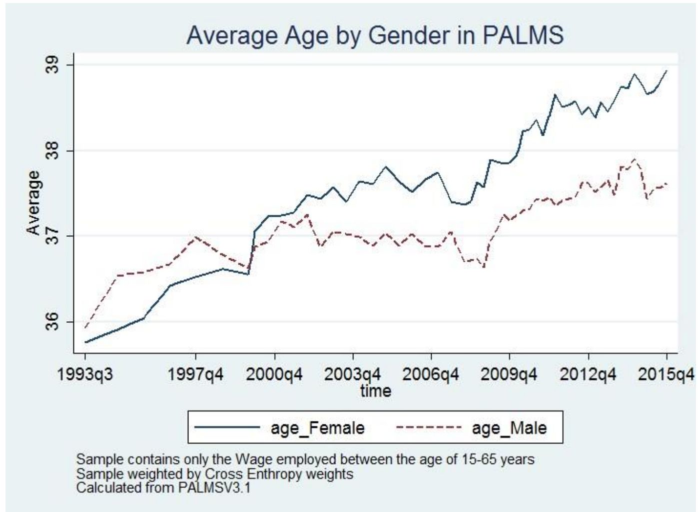


Figure 45: Average age by gender in PALMS

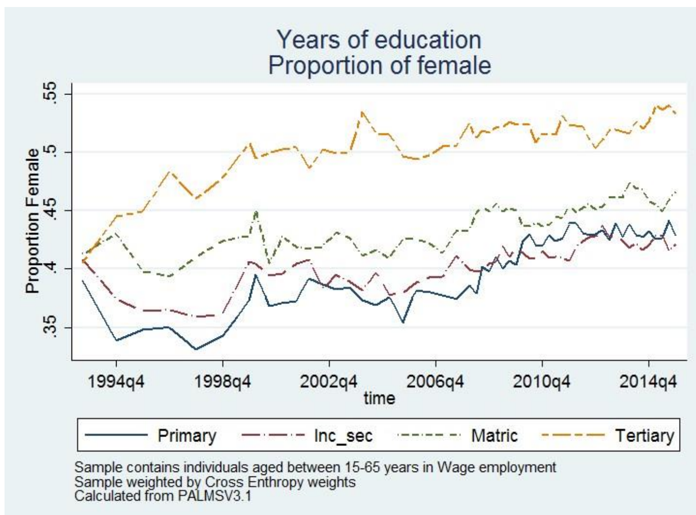


Figure 46: Proportion female: Education

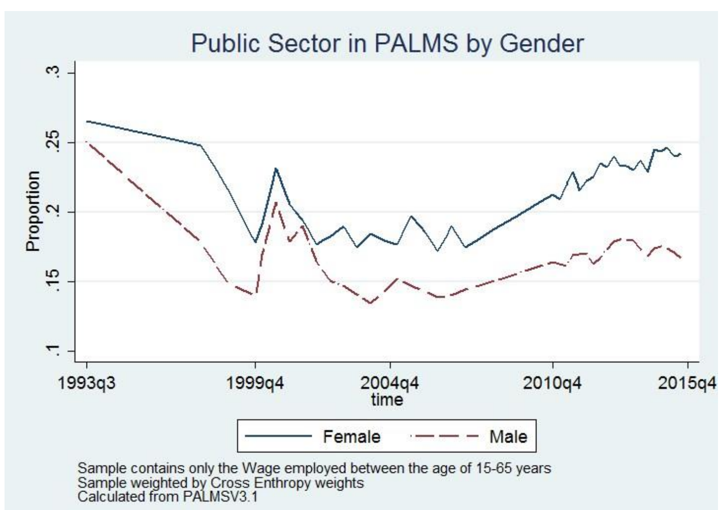


Figure 47: Proportion in Public Sector

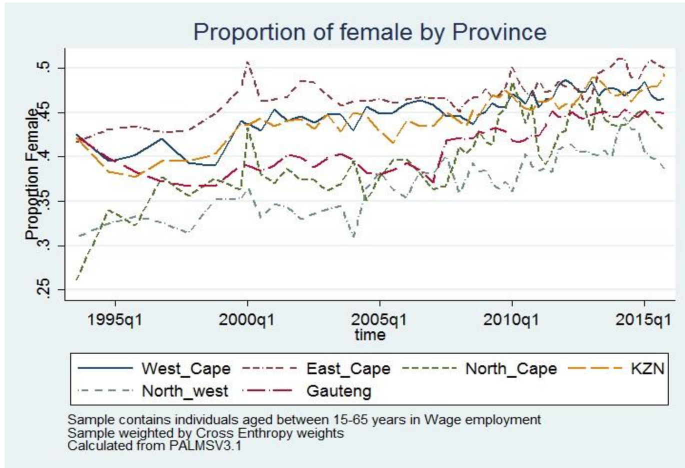


Figure 48: Proportion female- Province

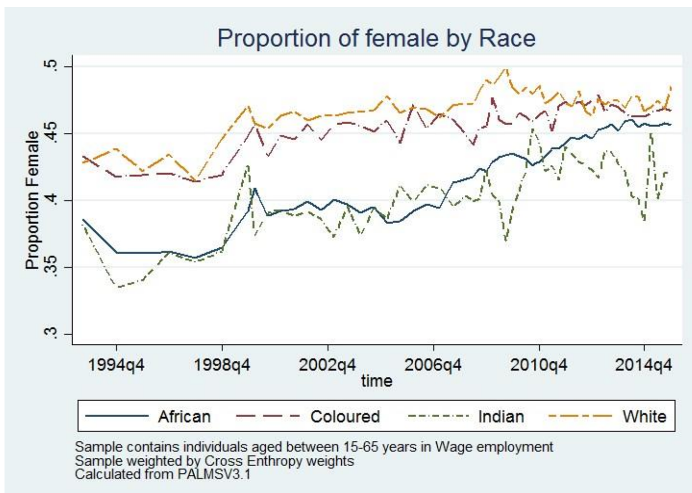


Figure 49: Proportion female- Race

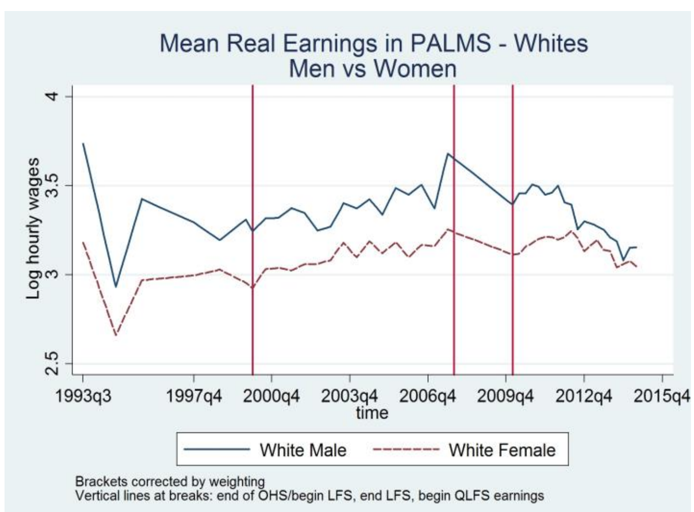


Figure 50: Wage series by gender- Whites

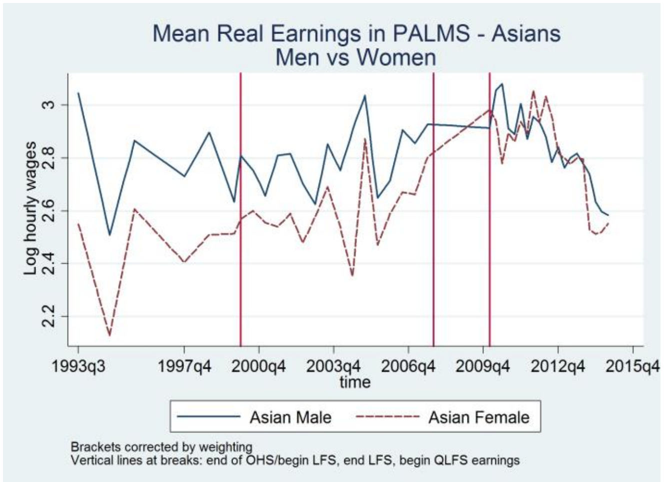


Figure 51: Wage series by gender- Asians

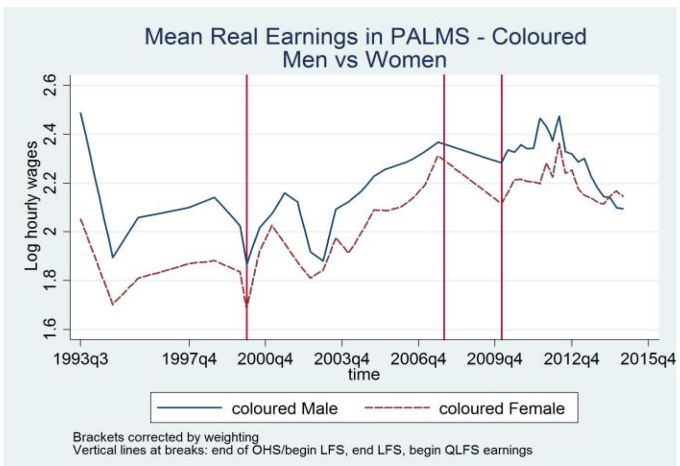


Figure 52: Wage series by gender- Coloured

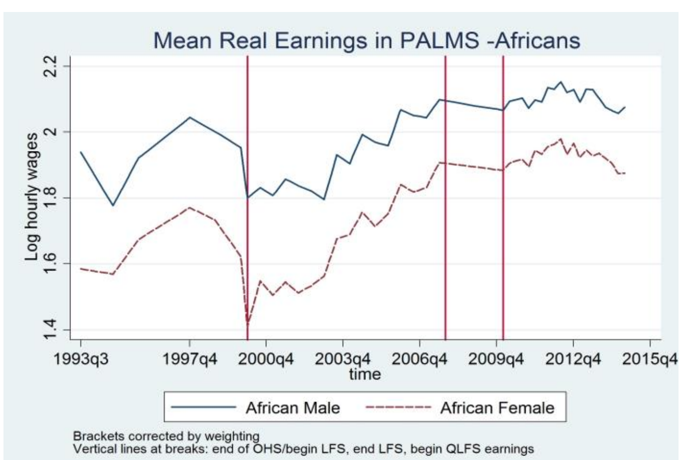


Figure 53: Wage series by gender- Africans

Table 15: Studies on the gender wage gap in South Africa

Table 15: Studies on the Gender wage gap in South Africa

Study	Data	Sample	Estimation Methods	Selection Bias	Controls	Finding
Winter 1999	OHS 1994	Formal sector workers	OLS, OB	No	Experience, Experience squared, Education, Log hours	female-male wage ratio All (0.87), Africans (1.01), Coloured (0.88)[a], Asian (0.80), whites 0.67
Hinks 2002	OHS1995	age 15-65 defined education	OLS, OB	Lambda insignificant	Age, age^2, Education Tenure, Occupation, province, union, sector	Gap Africans (-0.096), Coloureds (0.0455), Indians (0.24), whites (0.46)
Grun 2004	OHS1995,1997 and 1999	Africans and whites	OB	yes	Age, age^2, Education province, union, sector Occupation,union, income category rural, tenure,tenure^2	Africans:1995: (-0.1799), 1997: 0.1203, 1999:0.1294 Whites:1995: 0.5137, 1997: 0.3352, 1999:0.3350
Rospabe 2001	OHS 1999	age 15-65	OB, Interval Regressions	No	Race, Experience Tenure^2, Occup, prov, union, sector, Formal sector, Education Urban, Tenure, married	All (0.257), Africans (0.34) Coloureds (0.192), whites (0.349), Indians (0.202)
Ntuli 2007	OHS 1995,1999 LFS 2004	Africans Formal	quantile regressions	Lambda insignificant	Age, age^2, log hours Education province, union, sector Occupation married	Raw Wage gap in quantiles 1995: $\theta = 10$ (0.56), $\theta = 25$ (0.41), $\theta = 50$ (0.36), $\theta = 90$ (0.11) 1999: $\theta = 10$ (0.56), $\theta = 25$ (0.76) $\theta = 50$ (0.55),$\theta = 90$ (0.13) 2004: $\theta = 10$ (0.66),$\theta = 25$ (0.61) $\theta = 50$ (0.60), $\theta = 90$ (0.05)
Muller 2009	OHS 1995-1999 LFS 2001-2006	Part-time and Full-time workers	OB, JMP	No	Model I-Race, Age, Education, Marital status,Children Model II- Covariates in model I plus Tenure, Province, union, Firm size, sector, Occupation, Model III- covariates in model II plus Conditions of work that is, sector of employment Medical insurance, Pension, (Formal/Private)	1995 Full Time -0.020[a], -0.030[b] 1999 Full Time 0.239[b], 0.245[a] 2001 Full Time 0.209[c,a], 0.203[b] 2006 Full Time 0.172[a], 0.159[b] , 0.162[c]
Bhorat and Goga 2013	LFS 2007(Sept)	Africans	RIF	No	Experience, Experience^2, married, Province Education (5 categories), wage employed Formal sector, Occupation, pubic, Sector	10th Quantile gap 0.632 50th Quantile gap 0.35 90th Quantile 0.072

Note: a=Model I, b=Model II c=Model III
Source:Winter (1999, p.30 and table 16 p.31), Rospabé (2001, Table 3A P.36),Hinks (2002, Calculations from table 4 p.2048), Ntuli (2007b, Table 3 p.24), Bhorat & Goga (2013, table 3 col 1 p.15) Muller (2009, Table 7 and 8 p.22)

[a] In the text it says 88% but in the table it is written 0.80

APPENDIX B: CHAPTER FOUR

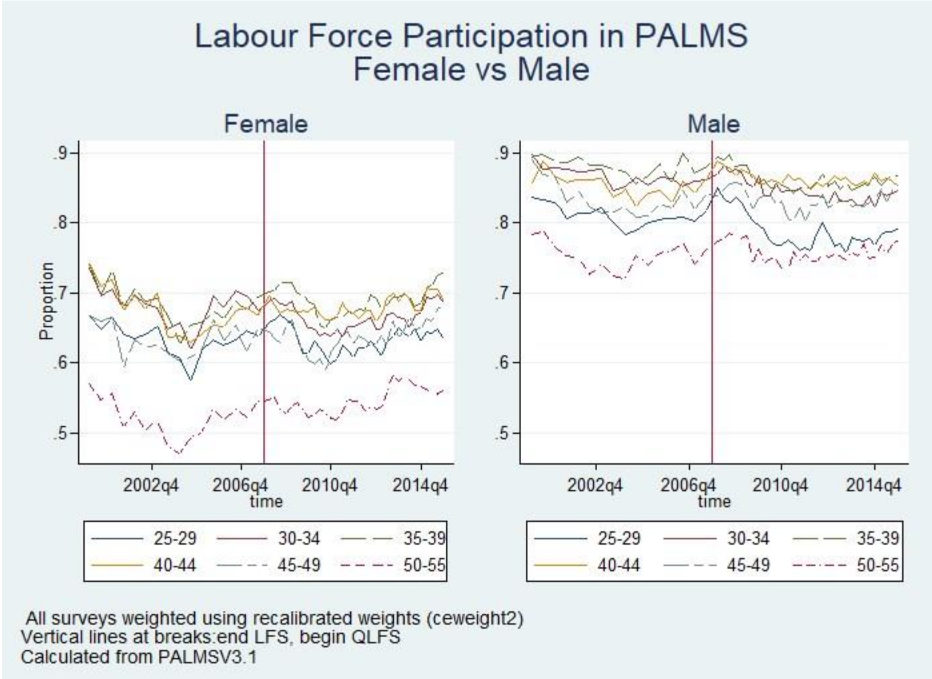


Figure 54: Labour Force Participation over time

Table 16:Age, Period and Cohort effects Regression output

VARIABLES	(Equation 21) Female	(Equation 22) Male	(Equation 23) Wage gap
Incomplete_sec	0.129	0.200	-0.00657
	(0.154)	(0.153)	(0.145)
Matric	0.861***	0.711***	-0.190
	(0.163)	(0.162)	(0.183)
Tertiary	1.840***	1.702***	-0.327**
	(0.161)	(0.181)	(0.158)
East Cape	0.0825	-0.485	-0.217
	(0.279)	(0.352)	(0.303)
Northern Cape	-1.001	-0.259	0.926
	(0.835)	(0.885)	(0.902)
Free State	-0.310	-0.709**	-0.254
	(0.306)	(0.355)	(0.328)
KwaZulu Natal	-0.100	-0.438	-0.0203
	(0.235)	(0.293)	(0.269)
Northwest	-0.282	-0.159	0.129
	(0.283)	(0.335)	(0.302)
Gauteng	0.0835	-0.149	-0.0172
	(0.212)	(0.278)	(0.259)

Table 16 continued

Mpumalanga	-0.0305	-0.0970	-0.11
	(0.285)	(0.341)	(0.326)
Limpopo	-0.518**	-0.388	0.512*
	(0.257)	(0.340)	(0.309)
married	0.0818	0.354***	-0.0308
	(0.120)	(0.117)	(0.0531)
Aged26	-0.00129	0.107***	0.132***
	(0.0325)	(0.0365)	(0.0397)
Aged27	0.0234	0.116***	0.141***
	(0.0359)	(0.0387)	(0.0469)
Aged28	0.0941**	0.164***	0.123***
	(0.0386)	(0.0394)	(0.0387)
Aged29	0.182***	0.173***	0.0689
	(0.0459)	(0.0408)	(0.0456)
Aged30	0.194***	0.257***	0.135***
	(0.0474)	(0.0424)	(0.0423)
Aged31	0.226***	0.211***	0.0868*
	(0.0412)	(0.0448)	(0.0450)
Aged32	0.312***	0.243***	0.0711
	(0.0425)	(0.0465)	(0.0434)
Aged33	0.329***	0.313***	0.141***
	(0.0437)	(0.0500)	(0.0493)
Age34	0.372***	0.319***	0.0740
	(0.0424)	(0.0506)	(0.0463)
Aged35	0.383***	0.338***	0.0890*
	(0.0425)	(0.0516)	(0.0477)
Aged36	0.410***	0.357***	0.146***
	(0.0507)	(0.0551)	(0.0548)
Aged37	0.475***	0.381***	0.0532
	(0.0471)	(0.0565)	(0.0504)
Aged38	0.520***	0.444***	0.111**
	(0.0522)	(0.0559)	(0.0524)
Aged39	0.538***	0.411***	0.0952*
	(0.0552)	(0.0613)	(0.0561)
Aged40	0.535***	0.440***	0.0888*
	(0.0506)	(0.0606)	(0.0523)
Aged41	0.532***	0.452***	0.207***
	(0.0542)	(0.0631)	(0.0553)
Aged42	0.532***	0.448***	0.202***
	(0.0493)	(0.0620)	(0.0557)
Aged43	0.584***	0.457***	0.0882
	(0.0526)	(0.0619)	(0.0567)
Aged44	0.608***	0.486***	0.161**
	(0.0542)	(0.0621)	(0.0656)

Table 16 continued

Aged45	0.641***	0.510***	0.145***
	(0.0497)	(0.0631)	(0.0527)
Aged46	0.630***	0.498***	0.157***
	(0.0539)	(0.0639)	(0.0601)
Aged47	0.666***	0.482***	0.146***
	(0.0506)	(0.0647)	(0.0561)
Aged48	0.654***	0.575***	0.187***
	(0.0512)	(0.0629)	(0.0550)
Aged49	0.646***	0.504***	0.127**
	(0.0490)	(0.0658)	(0.0610)
Aged50	0.746***	0.537***	0.0871
	(0.0577)	(0.0666)	(0.0601)
Born 1988	-0.0257	0.185	0.177***
	(0.0525)	(0.114)	(0.0502)
Born 1987	0.0203	0.0126	0.0372
	(0.0481)	(0.109)	(0.0960)
Born 1986	0.106*	-0.0152	-0.151**
	(0.0617)	(0.108)	(0.0723)
Born 1985	0.0816	-0.00728	-0.136*
	(0.0592)	(0.105)	(0.0726)
Born 1984	-0.0118	-0.0655	-0.110*
	(0.0513)	(0.107)	(0.0606)
Born 1983	0.000607	-0.0897	-0.104
	(0.0610)	(0.107)	(0.0643)
Born 1982	-0.0620	-0.154	-0.0833
	(0.0619)	(0.106)	(0.0618)
Born 1981	-0.0645	-0.131	-0.0817
	(0.0510)	(0.105)	(0.0596)
Born 1980	-0.176***	-0.167	0.0269
	(0.0595)	(0.105)	(0.0762)
Born 1979	-0.217***	-0.176*	0.0517
	(0.0549)	(0.105)	(0.0700)
Born 1978	-0.157**	-0.207*	0.0296
	(0.0625)	(0.105)	(0.0655)
Born 1977	-0.176***	-0.239**	0.0249
	(0.0585)	(0.107)	(0.0642)
Born 1976	-0.261***	-0.240**	0.100
	(0.0555)	(0.108)	(0.0671)
Born 1975	-0.247***	-0.257**	0.0386
	(0.0606)	(0.107)	(0.0658)
Born 1974	-0.241***	-0.268**	0.0504
	(0.0647)	(0.109)	(0.0741)
Born 1973	-0.282***	-0.281**	0.0805
	(0.0706)	(0.109)	(0.0686)

Table 16 continued

Born 1972	-0.284***	-0.270**	0.0675
	(0.0624)	(0.111)	(0.0776)
Born 1971	-0.222***	-0.311***	-0.0675
	(0.0688)	(0.112)	(0.0742)
Born 1970	-0.290***	-0.298***	0.0747
	(0.0700)	(0.114)	(0.0754)
Born 1969	-0.282***	-0.256**	0.0954
	(0.0676)	(0.116)	(0.0783)
Born 1968	-0.282***	-0.232*	0.0215
	(0.0782)	(0.118)	(0.0822)
Born 1967	-0.324***	-0.221*	0.172**
	(0.0753)	(0.117)	(0.0824)
Born 1966	-0.250***	-0.197*	0.128
	(0.0776)	(0.119)	(0.0876)
Born 1965	-0.357***	-0.221*	0.200**
	(0.0831)	(0.121)	(0.0926)
Born 1964	-0.291***	-0.233*	0.102
	(0.0781)	(0.123)	(0.0845)
Born 1963	-0.355***	-0.186	0.169*
	(0.0818)	(0.128)	(0.0916)
Born 1962	-0.322***	-0.215*	0.176*
	(0.0863)	(0.127)	(0.100)
Born 1961	-0.324***	-0.227*	0.181*
	(0.0832)	(0.132)	(0.0956)
Born 1960	-0.283***	-0.186	0.0899
	(0.0894)	(0.134)	(0.0973)
Born 1959	-0.342***	-0.174	0.0682
	(0.0956)	(0.139)	(0.111)
Born 1958	-0.355***	-0.290**	0.123
	(0.0964)	(0.137)	(0.103)
Born 1957	-0.348***	-0.127	0.228**
	(0.0951)	(0.139)	(0.111)
Born 1956	-0.323***	-0.159	0.131
	(0.0981)	(0.142)	(0.112)
Born 1955	-0.426***	-0.239*	0.174
	(0.111)	(0.145)	(0.109)
Born 1954	-0.388***	-0.252*	0.138
	(0.101)	(0.149)	(0.112)
Born 1953	-0.434***	-0.227	0.194*
	(0.110)	(0.150)	(0.114)
Born 1952	-0.528***	-0.338**	0.200*
	(0.117)	(0.156)	(0.120)
Born 1951	-0.410***	-0.252	0.152
	(0.109)	(0.168)	(0.125)

Table 16 continued

Born 1950	-0.500***	-0.401**	0.179
	(0.107)	(0.175)	(0.121)
Year2002	-0.105***	-0.0961***	0.0162
	(0.0207)	(0.0164)	(0.0166)
Year2003	0.00580	0.00472	-0.0135
	(0.0181)	(0.0168)	(0.0190)
Year2004	0.0473***	0.0380**	-0.0149
	(0.0181)	(0.0178)	(0.0232)
Year2005	0.0166	0.00752	-0.0317*
	(0.0185)	(0.0179)	(0.0172)
Year2006	0.0619***	0.0617***	0.000811
	(0.0173)	(0.0174)	(0.0218)
Year2007	0.122***	0.101***	-0.0441**
	(0.0200)	(0.0184)	(0.0200)
Year2010	0.0355**	0.0258	0.00151
	(0.0163)	(0.0176)	(0.0186)
Year2011	0.0194	0.00343	-0.0339**
	(0.0164)	(0.0167)	(0.0170)
Year2012	-0.00154	-0.0102	0.00878
	(0.0163)	(0.0165)	(0.0175)
Year2013	-0.0262*	-0.0203	0.0358**
	(0.0148)	(0.0160)	(0.0149)
Year2014	-0.119***	-0.0820***	0.0249
	(0.0165)	(0.0157)	(0.0193)
Constant	1.133***	1.553***	0.166
	(0.236)	(0.323)	(0.283)
Observations	338	338	676
R-squared	0.861	0.859	0.497

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable for the female and wage regressions as per equation 21 and 22 is the log of hourly wages while the dependent variable for the wage gap regression as per equation 23 is calculated as $logmalewage - logfemalewage$

Table 17: African Male

Table 17: African Male

Cohorts	Survey Year												
	2000	2001	2002	2003	2004	2005	2006	2007	2010	2011	2012	2013	2014
1950	173												
1951	97	109											
1952	200	129	143										
1953	163	159	119	119									
1954	155	146	135	108	109								
1955	222	149	121	115	94	112							
1956	148	177	111	117	115	126	126						
1957	193	135	145	107	93	134	110	110					
1958	262	197	149	169	124	122	132	145					
1959	173	217	163	114	159	120	116	118					
1960	262	189	231	221	136	184	144	132	119				
1961	210	204	173	196	136	104	142	117	95	98			
1962	293	230	229	149	198	157	133	147	130	139	147		
1963	259	256	194	204	125	186	148	132	141	135	129	109	
1964	265	221	205	194	169	141	187	135	135	133	127	135	108
1965	281	241	167	203	162	184	145	183	145	135	120	127	109
1966	215	205	162	174	166	154	187	126	128	147	126	146	97
1967	218	199	193	153	146	152	168	169	150	132	101	138	111
1968	306	187	191	225	175	174	186	175	163	149	143	139	119
1969	250	277	175	180	207	198	198	170	189	165	143	123	135
1970	321	249	229	197	226	267	234	209	182	179	183	155	169
1971	209	247	158	197	176	183	188	192	146	142	147	158	121
1972	264	218	213	186	237	197	202	231	153	160	156	150	139
1973	238	216	176	206	178	263	150	206	189	178	178	168	150
1974	248	207	200	194	217	180	246	200	176	157	200	171	172
1975	221	186	187	185	204	218	207	220	197	196	176	187	150
1976		181	156	169	194	216	227	196	185	196	187	180	156
1977			190	162	219	208	235	226	183	209	187	168	152
1978				154	154	170	203	229	190	179	193	193	165
1979					158	174	197	201	194	203	211	211	172
1980						170	237	240	209	216	232	204	182
1981							208	199	189	210	182	223	190
1982								227	209	231	208	212	191
1983									218	241	199	213	213
1984									205	218	208	230	204
1985									184	215	218	233	216
1986										173	201	212	200
1987											155	197	207
1988												183	183
1989													175

Source: Own Calculation from PALMS V3.1

Table 18: African female

Table 18: African Female

Cohorts	Survey Year												
	2000	2001	2002	2003	2004	2005	2006	2007	2010	2011	2012	2013	2014
1950	123												
1951	94	106											
1952	160	110	126										
1953	127	115	119	114									
1954	141	108	141	103	84								
1955	176	135	106	96	121	92							
1956	135	118	105	92	117	113	113						
1957	162	110	114	94	80	96	116	105					
1958	211	136	123	105	120	111	125	116					
1959	158	193	138	105	134	121	123	134					
1960	240	155	196	148	122	145	135	135	137				
1961	143	198	105	160	131	106	113	127	139	116			
1962	233	171	176	123	178	144	153	157	117	146	124		
1963	181	179	130	175	115	170	142	115	141	167	131	133	
1964	222	160	148	154	208	127	156	146	146	172	146	152	160
1965	216	184	126	145	126	165	128	170	139	172	163	163	141
1966	192	169	147	133	158	129	179	119	163	129	146	138	140
1967	166	152	142	131	137	162	145	153	161	173	152	134	122
1968	199	179	144	167	162	132	172	139	151	150	157	163	137
1969	160	174	110	103	151	150	133	154	173	182	165	158	175
1970	202	147	189	121	152	162	163	153	182	153	178	174	156
1971	149	184	113	150	138	131	145	152	128	146	165	151	126
1972	173	150	139	107	175	141	182	174	195	185	146	161	148
1973	165	157	119	145	119	156	144	148	155	173	173	156	154
1974	148	125	111	111	175	136	166	152	161	138	183	179	151
1975	141	124	122	130	131	166	139	165	165	171	145	186	168
1976		100	99	104	131	122	175	124	168	165	158	157	145
1977			124	97	131	155	147	172	142	161	164	154	156
1978				92	109	132	142	156	177	152	166	143	154
1979					119	121	118	151	159	160	161	163	136
1980						117	133	159	143	166	189	166	162
1981							129	118	143	149	154	190	160
1982								136	166	153	163	150	175
1983									174	165	195	183	176
1984									137	173	158	154	173
1985									137	143	172	177	178
1986										135	138	170	160
1987											144	158	166
1988												144	136
1989													136

Source: Own Calculation from PALMS V3.1